A determination and classification of student errors in lower-level calculus through computer-aided assessment and analysis

A thesis submitted for the degree of Master of Philosophy

By

James Hanson

Department of Mathematics, Brunel University, UK

March 2011

Abstract

This thesis describes the determination of student errors through paper-based assessment through to computer-aided assessment. The focus on identification of errors is through low-level calculus questions; on polynomial differentiation through to product rule and chain rule usage in integration.

A major objective of this work is the design of suitable computer-aided assessment to fulfil learning objectives and usage as a diagnostic tool. The limitations of such diagnostic tools (whether paper-based or computer-aided) are explored in some depth, and the role of question design is brought to the fore through careful analysis of student errors on summative as well as formative assessments. It is from the data gained through these assessments that we can begin to classify mistakes into usable taxonomies. Firstly I have chosen to use the ResultsPlus data files from summative assessments via Edexcel, secondly the results from several years of paper-based formative assessment on a multiple-choice diagnostic test, and thirdly formative testing scores from Brunel University first year Economics students. The basis for forming an over-arching taxonomy for mistakes is built up using the SOLO model for classification and the discussion turns very much back towards question design as the nature of student errors changes as question structure changes. The data generated through computer-aided assessment is firstly unpacked to allow comparison between difficulty levels and cognitive levels. I go on to look at temporal comparisons between cohorts over time to discover weaker skill areas and question discrimination that will yield improved diagnostic assessments as well as selectively difficult assessments for the top-end of a cohort. The thesis looks carefully at the limitations of classification from these data sets and explores further distracter design in computer-aided assessment questions.

Acknowledgement

I acknowledge the department of Mathematics at Brunel University for accepting my application to join them as a part-time postgraduate research student, and express my thanks to Dr. Martin Greenhow for his support and guidance throughout.

I declare that this thesis I hereby submit for the degree of MPhil at Brunel University, UK, is my own work, supervised by Dr Martin Greenhow. I have not previously submitted the work at any university or for any other publication.

James Hanson
Department of Mathematics
Brunel University, UK

Foreword

This thesis is primarily for teachers of both students preparing for the transition between GCSE and GCE maths, as well as those studying for AS maths. It will also be useful for those teaching remedial maths courses / foundation maths courses at tertiary level who have only GCSE maths or basic understanding / competence at GCE maths. The major aim is to enable teachers to use CAA in assessment, and help students use CAL to boost understanding. A secondary aim is to help teachers and students alike in diagnosis of common errors and areas of weakness or misunderstanding so that they can make great strides in learning.

I have spent ten years full-time teaching maths to secondary level students, at two different secondary schools. Before that, after my initial maths degree, I completed an MSc in mathematical biology, seeking to pursue a relevant interest in medicine and applications of mathematics to medicine. While teaching, I have been fascinated to see how students make the transition from GCSE to GCE maths, how the gap has narrowed from syllabus changes, examination material release, changes to assessment strategies and changes to incorporate IGCSE. My research journey led to me to use a diagnostic test to examine bridging materials in core maths skills I perceived vital for future GCE success. I also embarked on a collection of data through a basic paper-based test, and this in turn led me to the professionals at Brunel and Martin Greenhow, looking at Mathletics. I started by writing a few questions for Mathletics on surds and indices for the package, mainly to facilitate my own understanding of question pedagogy, the use of CAA and CAL and also the role that CAA could possibly play in the school curricula. From here, my research interest turned more to the pedagogy of writing questions for my own students to use repeatedly as CAL, and also as paper-based diagnostic tests. I wanted to really investigate how robust present-day CAA was for students, and how close we are to designing reliable assessment models that take mathematics away from paper to onscreen.

James Hanson

March 2011

<u>Contents</u>

List of Tables, Figures and Graphs

Chapter 1: - Introduction

1.1 Role of Assessment in the school curriculum

In the UK, the tradition for assessment has been for terminal public exams (A levels, Highers, and national curriculum tests etc) to assess the (centrally agreed) syllabus that is taught in schools, **Baker** (2010). This is much less true in other countries where the exams are less closely tied to the school curriculum.

The system of grading for each examination used to be such that only a set proportion of students passed each year, in order to filter out the (then) small percentage that were deemed suitable for entry to tertiary education. Clearly it would have been impossible to judge how standards changed over decades, or even how much students knew at any one year, since the removal of such bars in the 1980s.

Summative assessment in mathematics has changed little in 100 years, and the University of London schools matriculation examination, **Carson & Dale** (1909) is one such exam paper that I have furnished my students with this year, as a means of enriching them with a different style of questions which still rely on skills present in today’s curriculum. In fact, many of them find this assessment more interesting than standard present-day assessments.

It could be argued that this staleness in approach has robbed students of potential gains in achievement, **Emanuel** (2010), and that those who are taught only to pass the test, rather than encouraged to learn skills, leave school bored by endless testing, **Ellis** (2010). While teachers often teach beyond the requirements of any examination, there is widespread acceptance that “what you assess is what you get”, **MEI** (2010), and students are widely asking “is this going to be tested? And if not, then why should I learn this?”.

The argument for a change has begun in the past decade, not only with the advent of technologies sympathetic to learning and assessing, but also the growing unhappiness with the status quo in teaching purely to a summative written test.

1.2 Teaching to the test

The traditional assessment model has always been where students' responses to individual questions and sub-questions are assigned marks to 'correct' responses, and marks lead to grades. The problem with this is that it encourages students to play the assessment game, and go for a mark 'trawl' in exams, trying to pick-up bits of marks here and there. It also allows students to retake modules and units of assessments, in order to gain a higher mark and grade, often encouraged by schools. The economics of such an approach must be a heavy counter-argument before you look at the efficacy of this assessment strategy. The alternative is criteria-based assessment where grades are awarded according to how well students meet the desired learning outcomes, **Biggs (1999).** In an ideal world, we would like to reward students on how well they have achieved such individual learning outcomes, pre-specified before the course started, but that approach is enormously difficult to get right or fair.

Some 27% of entries (to A-levels) gained an A or A* and the overall pass rate rose for the 28th year, amid a record battle for university places, **Harrison** (2010). Many questions pertaining to rising pass rates or decreasing standards are asked every August when exam boards and schools publish seemingly improving results in A-levels and GCSEs to hundreds of thousands of teenagers across the country. Authorities now spend large amounts of time and effort trying to locate the 'easiest' exam boards for boosting their schools' results, **Golding** (2010). It is clearly of great distress to students who achieve high results, to hear the national media bashing their achievements, and forever comparing the assessment that they have just sat to that which was sat a decade or more previously.

What we need is the ability to accurately compare assessment performances on a temporal basis, but there is little consensus about how to compare standards, definitively, over long periods of time, **Baker** (2010). I, for one, would like to be able to compare results achieved in 2011 with those from 2001, and indeed 1991, so see how students of my generation compare to those I teach today.

Those arguing that media-based allegations of “grade inflation” are unfair seem to be lone voices in the assessment world, and often their impassioned defence of the consistent standards falls on deaf ears. However **Murphy et al.** (2010) are in agreement that exam papers (across the school curricula, including maths) have to change, for the simple reason that the curriculum changes all the time, and assessment (which produces exam grades) is not an exact science, so cannot be accurate enough to conclude that standards are changing at all. While there is little concrete evidence backing up their confidence in our consistent system of assessment, the there is growing evidence of degradation in the quality of A level grades with real mathematical knowledge of students becoming increasingly porous, **Barry et al.** (2003).

The fickle nature of examination success is clearly intertwined with entry numbers, and there is evidence of significant entry increases in STEM subjects, long rejected by students and grade hungry schools in favour of “soft” A level subjects. These entry increases are coincident with ever increasing pass rates and A grade proportions in A level mathematics students, **Harrison** (2010). This increase in pass rates is clearly not down to the fact that students have worked harder than ever, nor that standards overall continue to improve, because universally employers and universities are complaining that entrants appear to lack even basic numeracy skills, **Spriggs** (2010). At present, schools are generally thinking of how to maximise league table positions, A level grades, and not what A levels are valued by universities, **Ricketts** (2010).

This hard position has been long understood by those outside of the education world, but it is only recently that those within education have seen what is wrong with the assessment system of today. The fact that exam papers have become very predictable, **Smithers** (2010), and that there is competition for lucrative custom, **Golding** (2010) all point to the same inextricable conclusion: assessment rewards those who pursue ‘teaching to the test’. Even teachers feel obliged to buy the latest textbooks and teaching resources endorsed by the exam boards, fearing that their pupils would be disadvantaged if they did not **IMA**, (2010).

Summative assessment, in its present form, has to change, and learning outside of such narrow channels must be the major benefit. So the learning process needs formative assessment / diagnostic tests of high quality. How they can be generated in a consistent, robust, repeatable, evidence-based, defensible way is the subject of this thesis.

1.3 Research Questions and structure of the thesis

I found myself instinctively drawn to (perhaps neutral) research questions, as a desire to test how far we could go within school curricula, without a firm hypothesis when I started the research project.

1. Do students make the same mistakes online as on paper-based tests?
2. Can we devise a diagnostic test spanning core maths skills?

The ultimate question for the future of Computer-aided assessment (CAA) – can we reliably tell if the nature of assessment will radically change the mistakes students will make and will it reliably test the range of skills we wish to assess using it as our assessment vehicle?. Given evidence showing that CAA is likely to grow within the school curricula, we must have diagnostic and formative tests of high quality. My hypothesis here is that students don’t make the same mistakes on paper as online, as I believe the nature of the question, its type and required response change the way of approach greatly, and as such the model for assessment must move its aims and quality to be able to test a wide range of skills in the same way and also generate reliable results.

3. Can we identify or infer and classify student mistakes using evidence from students' work?
4. Can we generate a taxonomy for student errors?

My hypothesis here is that we can construct a basic taxonomy for classifying student errors, but it is beyond the scope of my ability to construct a fully detailed tree-like structure taxonomy to cover the multitude of student errors. In short, I wanted to propose that we could generate a fully automated question taxonomy model that would classify every mistake for us. But this has proved impossible, and I conclude that the human question author is needed. I wanted to see how far we could go in classification, and propose a model that reflects the mechanical nature of questions. This in itself has a general use in construction of diagnostic and assessment questions because we can (at least on a question by question basis) classify errors using careful question structures. The research was also focussed on seeing how much we can infer from students' results files, what we can conclude about present day paper-based tests and classification of errors.

5. Can we generate a reliable model for choosing and refining mal-rules for diagnostic calculus questions?
6. Can we understand how the choice of mal-rules affects the difficulty of multiple-choice objective questions on calculus?
7. Can we use students' known mistakes and our own evidence bases to improve presently used questions on calculus?

Calculus is a wide topic, but was selected as a narrowed down field of research, as there is overlap between some IGCSE (International GCSE) courses and the start of the GCE course, it underpins all core maths modules at GCE level, and the core maths 1 skill pyramid has calculus underpinned by many other key bridging topics such as indices, algebra, graphs, polynomials. There are also a large number of questions written on calculus within many CAA and Computer-aided learning (CAL) products, as it could anecdotally be accepted that understanding of calculus is broadly indicative of general maths understanding. The subject of mal-rules is one I have developed a really good understanding of regarding the pedagogy, I wanted to investigate how mal-rules can give false-positives, or change the scope of a question from facility to discrimination indices, so these questions were posed with an investigative slant. The hypothesis I wished to explore was that mal-rules written for questions can lead the students either towards or away from the correct answer a disproportionate number of times. I wanted to also test whether such mal-rules served as a very useful diagnostic tool, and whether we could have a high degree of confidence in both the nature of student errors using them, this would help generate our taxonomy of errors.

It is worth setting out at this stage the fundamental differences between the on-line and paper based forms of assessment (a fuller discussion of online testing and objective testing is to be found in section 1.5). Some of the differences are:

- On-line assessments are quite often made up from multiple-choice questions, whereas on paper-based assessment they are usually not made up in this way.
- In some cases there are multiple goes at the assessment on-line (with the highest score counting). On paper assessments it requires a modular course structure to allow such additional goes.
- On traditional assessment, you only require paper and pen, whereas on-line assessment, you require those and other resources (including computer labs and invigilation for controlled condition assessment).
- On-line assessment can give instantaneous marking and feedback, compared with longer delays for paper-based assessment.

We also need to distinguish between objective and subjective questions and which lend themselves to CAA. Most questions in maths are objective (meaning that they are readily assessed online), but proof, modelling (e.g. mechanics or sketch the graph of a function) and interpretation of results (e.g. in statistics questions) is subjective and is beyond the scope of online objective formats, however delivered. The closest we can usually get is to identify an error in a few lines of solution or proof (often defined as “Hot Line questions”, where students identify the wrong “Hot” line in a false proof). Even questions where the student has to choose between “true/false/undecidable” statements is merely asking students to respond not generate the proof/statement for themselves (unprompted).

Then we have to consider the way a CAA system has to communicate with its marking scheme. Humans are very flexible, computers not, so CAA marking needs even more clear guidance. So this usually means that the question generally has to be heavily structured or of multiple-choice format (MCQ). Even with a Numerical Input (NI) question the student will know that the answer is a number (possibly not a whole one if it specifies the number of decimal places required) and not say a matrix or an algebraic expression. For example, an integral with limits gives a number not an indefinite integral; the determinant of a matrix is a number not a matrix. Two input boxes might suggest there are two roots to be found, not just one etc, so we have to be careful how we lead students thinking.

Finally issues of partial credit need addressing for incomplete or inaccurate solutions; this can be programmed but a human would respond automatically to a solution that is almost correct by giving at least some marks. For a CAA system all we know is what was selected or typed in the answer box.

Within the thesis, I set out the main methodology in chapter 2. The basis for statistical analysis and choice of technology platforms, a glossary of testing terms as well as question pedagogy and question type are all justified here and chapter 1 as well. This chapter brings in the rationale for using multiple-choice questions in objective testing, and highlights the reasons for using diagnostic testing as well as outlining the key features of such tests. The first evidential chapter (3) shows how we identify student errors from their own work, firstly using extensive written feedback, and observed work at different levels and across different schools / colleges. This leads on to the development of a taxonomy of student errors in chapter 4, where I explore different learning styles, objectives and outcomes, and finally demonstrate which taxonomy of errors I propose to use throughout the thesis. I use the end of the chapter to show how the body of evidence from chapter 3 can be classified in this way, and how likely future mistakes on calculus questions could already be inferred from these results without the need for much more evidence.

In chapter 5, I aim to set up a diagnostic test using an algebraic framework, with previous paper-based non-objective tests used to generate common student errors, and show how this can be written into a new multiple-choice objective test. I investigate not only how efficient this test is at diagnosis, but also the issues with encoding common mistakes into such tests. I finish the chapter by drawing on the limitations of this model based either on paper tests or onscreen tests using Excel as our framework. Chapter 6 is the second main evidential chapter, which embarks on a thorough analysis of 374 students' exam scripts, with the aim of diagnosing profile errors by topic and skill levels, as well as trying to infer more specific errors on questions. Chapters 7 and 8 use thousands of online maths diagnostic test question responses from Brunel University's 1st year Economics students on differentiation and integration. The analysis from these chapters centres on classifying questions by difficulty / complexity and cognitive levels and comparing and contrasting between groups and cognitive levels. The chapters build on specific student errors, and a fuller analysis of the quality of the questions in their present form, particularly the efficacy of choice of mistakes to encode into the questions. I finish the thesis in chapter 9 with an over-arching conclusion and discussion on possible future developments in diagnostic testing, particularly using the computer-based assessment framework.

1.4 Learning or teaching strategies to avoid errors

Alongside these latest resources supplied by exam boards, more and more teacher support courses are offered that help guide teachers to cover all of the crucial points a chief examiner wants to see, with a minimum of time wasted on "extraneous" material, **Emanuel** (2010). A successful teacher, these days, seems to be one whose students have achieved the highest possible mark in the GCSE or A-Level examination without gaining a single skill or piece of knowledge outside of a national curriculum. Thus it could be argued also that the student who listens best to the teacher on the subject of what will come up and how to avoid known common errors in the nearly always predictable assessments is judged the brightest in the class, rather than the inquisitive one who is better at problem solving by themselves.

This digital age is a golden one for the development of technologies to support learning. It would be the worst possible moment to abandon the use of technology within learning, but exactly the right moment to restart the discussions about how to further its implementation, **Puttnam** (2011). Teachers themselves also need to undergo training in using such technologies, as no education system can be better than that enabled by the quality of the teachers.

However, there is significant division within government on the role of technology in education, with plenty who would leave the nature of the classroom as it is, in clear opposition to those who want to march education and assessment into the digital age. Speaking with those who work in the field of assessment, I sense that within a decade we will see widespread, if not wholesale computer-aided assessment or other technology platform-aided assessment. There is already a plethora of technology-aided learning resources available to students and teachers alike, **Drought** (2010). The thrust of this thesis will examine further how we can embed technology in the development of learning resources, diagnostic tools and teaching assistance to develop skills, equip students for higher learning and reduce the likelihood of errors in summative assessment.

1.5 Objective tests v non-Objective tests

Objective tests require a user to choose or provide a response to a question whose correct answer is predetermined. The converse is that non-objective tests are by nature much more subjective and often involve presenting and defending a viewpoint. Grades often reward the quality of the arguments presented rather than the position being defended.

Objective questions generally fall into a number of different types (multiple-choice, selection of true / false, matching correct statements or answers from a list, identifying objects or their positions or supplying a numerical or textual response). As the correct answers to objective test questions are pre-determined, they are well suited to the many forms of technology platform that involve automated marking. The electronic marking of the responses is completely non-subjective as no judgement has to be made on the correctness or otherwise of an answer at the time of marking. However, it is worth noting that in terms of in-built bias, an objective test is only as objective as the test's designer makes it, **CAA** (2002), for example in writing tests for my own students, my objectivity only tends to span the perceived ability range of the cohort.

The design of such tests, and the selection of the question type is crucial in making sure that assessment via objective tests will provide enough accurate information to classify students' understanding. Mathematics assessment lends itself better than most subjects to objective tests, as many skills can be tested using questions that require a numerical response, or a selection from a list of suitable answers. A popular type of question for the objective test developer in mathematics tests is the multiple-choice question (MCQ), whereby a number of similar possible responses can be drafted for each question. The further development of these questions is made possible by understanding common errors in students' understanding of individual skills, and by writing some of the possible responses to reflect these same errors (mal-rules), a pedagogy explained by **Greenhow et al**. (2003). However, its limitations will be discussed further throughout this thesis, as the whole areas of proof, modelling and interpretation in mathematics are ones that are difficult, but not impossible, to bring in to objective testing.

1.6 Formative Assessment

Whereas Summative Assessment involves measuring what has been learned in formal assessment, Formative Assessment in its widest sense refers to any process by which pupils are made aware of how they can make progress, **Clarke** (2010). Teachers have always been using formative assessment in the classroom (often via paper-based tests, or marking of class / home work) to enhance pupil learning, and such assessment should be by nature informative to the pupils (through feedback), and quantitative so that there is a measurable pattern of learning records for each pupil throughout a given timeframe. Many electronic platforms have embedded formative assessment strategies into their learning resources (e-learning), for example in regular bite-sized end of unit tests of skills.

Such assessment strategies, as part of e-learning resources have become more and more the norm for schools, as students are ever more technology hungry and expectant.

However the key questions still remain:

- How can we use these formative assessments to guide learning?
- How well will students use the feedback provided onscreen?
- How can this prepare students for summative assessments?

Black & Wiliam (2010) postulated that when formative assessment is most successful, students learn effectively, teachers focus more individually on students' needs, wider learning is readily enhanced and even the least able benefit as they feel involved in directing their own learning. Thus there is plenty of scope for development of wide ranging formative assessment tools embedded into other learning resources for students to use through the formal school setting as well as through independent learning time. The development of specific tailor-made resources online and through many differing technology platforms (e-resources) will only enhance technology-driven learning (and in the main that means computer aided learning - CAL).

Feedback needs to be given as soon after a formative assessment has been completed to be most effective, and should give students a sense of what has been understood as well as what is yet to be grasped fully. Marks and grades can be very empowering to some students, but can de-motivate students to the point that they neglect to take up the specific feedback. **Clarke** (2010) also advocates the use of oral feedback (including discussion) as the most effective type of feedback, as well as encouraging students to reflect on the feedback and given time to work on improvements as soon as possible after the assessment.

Preparation for summative assessment will only be enhanced if the formative assessments accurately mirror the summative ones, provide accurate feedback on specific skill deficiencies or gaps in knowledge and are used regularly enough to train the students to get the maximum from that style of learning. In short, the e-resources will be of tremendous help when summative assessment is conducted via e-assessment, and when these tools allow for an accurate diagnosis of ability, knowledge, and skill development. Until such time (perhaps a decade or so in many subjects), then they just serve a purpose of enhancing wider learning and providing useful additional backup to traditional teaching and learning strategies.

1.7 Question types: by pedagogy and by cognitive level

Question types used in this thesis are those specified by **Greenhow (1996)**, who divides assessment questions into five types, three of which I describe below as present during our analysis of computer-aided assessment objective tests:

- **Multiple-choice questions** (MCQ) have been highly used and are well regarded objective questions, **Harper (2003)**. They are built up from a question stem, and a set of options given including the correct answer (the key) and a number of associated distracters, so that students check the option box onscreen (or input an associated key coded letter). A good MCQ should contain plausible, but definitely wrong options as its distracters, **Baruah (2007)**. The emphasis on research through this thesis is on derivation and formulation of good distracters. They are highly useful for diagnosis, allow ease of marking and delivery of tailored feedback and students find them user-friendly. However, they aren't representative of usual paper-based assessment forms, and students can either be led to guess a solution, or be disappointed if their answer doesn't appear in the list (even though "None of these" could be correct).

- **Numerical Input questions** (NI) are often used to test a skill(s) whereby the MCQ might give away too many implicit cues for the students or "scaffold" their thinking. These questions ask for one or more numerical responses in a provided box or on paper. Such questions need to be carefully expressed onscreen and on paper to specify accuracy required, units required, and objective questions of this form are used most often to look at practical applications of laws / rules, rather than the recall of laws / rules. These questions are probably the easiest to formulate and write for on-line assessment, are familiar to students, relatively easy to complete on-line and quick to mark, but they don't allow easy scope for writing algebraic answers, and can't give specific diagnosis of errors or tailored feedback.

- **Responsive Numerical Input questions** (RNI) are identical in presentation to students as those of NI questions. However, the question is encoded with pre-prepared mistakes (hidden distracters), so that the feedback to students can provide more tailor-made explanations of where they have gone wrong, alongside the specific feedback about the full correct working to obtain the correct response. Such questions allow the collection of more specific responses from students, to ascertain the frequency and type of mistakes made. These serve as useful questions to derive distracters for future similar MCQs. They are highly useful for diagnosis, allow ease of marking and delivery of tailored feedback and students find them user-friendly. Their one weakness is that they don't allow easy scope for writing algebraic answers.

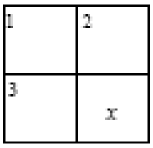
Question types can also have metadata for cognitive level or learning objectives, as well as by the categorisation of mistakes and number of structures that they are likely to require to complete. Such fuller descriptions will be described at the start of chapter 4 onwards.

1.8 Use of multiple-choice questions in formative and summative assessments

Multiple-choice questions are becoming more and more common through formative and summative assessment modes, and students are very much more aware of how they can be used in assessment and learning. From an early age, many secondary school level students will meet the United Kingdom Mathematics Trust (www.ukmt.org.uk) challenge papers, from Junior level (UK school years 7-9), Intermediate level (UK school years 9-11) and Senior level (UK school years 9-13). There are several stark observations with the nature of this different assessment through only multiple-choice questions (as shown in figure 1.1 below):

- The papers are strictly controlled so that candidates have no access to calculators, rulers or other measuring implements
- The papers are very much designed with time constraints that make them difficult to fully complete every question
- On some sections of the papers, there are negative marks available for wrong answers, zero marks for blank answers and 4, 5, or 6 marks available for correct answers.
- Answers are designed so that guesswork alone should not be readily possible, and a strategy of systematic elimination of answers by inspection should also not be successful.

1. What is the digit x in this cross-number?



Across — Down

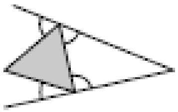
1. A cube — 1. One less than a cube
3. A cube

A 2 B 3 C 4 D 5 E 6

2. What is the smallest possible value of $20p + 10q + r$ when p, q and r are *different* positive integers?

A 31 B 43 C 53 D 63 E 2010

3. The diagram shows an equilateral triangle touching two straight lines.
What is the sum of the four marked angles?



A 120° B 180° C 240° D 300° E 360°

4. The year 2010 is one in which the sum of its digits is a factor of the year itself.
How many more years will it be before this is next the case?

A 3 B 6 C 9 D 12 E 15

5. A notice on Morecambe promenade reads: 'It would take 20 million years to fill Morecambe Bay from a bath tap.' Assuming that the flow from the bath tap is 6 litres a minute, what does the notice imply is the approximate capacity of Morecambe Bay in litres?

A 6×10^{10} B 6×10^{11} C 6×10^{12} D 6×10^{13} E 6×10^{14}

Figure 1.1: A screenshot from Q1-5 of the UKMT Senior Maths Challenge paper 2010

The other areas where students will increasingly come across multiple-choice questions are for GCSE revision exercises onscreen. Exam boards (as well as e-learning websites such as the University of Cambridge's www.nrich.org.uk site, the MyMaths team at www.mymaths.co.uk, or the BBC bitesize revision guides at www.bbc.co.uk/bitesize) are increasingly turning to use them as a diagnostic and formative assessment tool, so teachers can use them in classes within a computer room, or students can have independent access to the site so to practise on their own, and use the plethora of associated e-learning resources as spin-offs. Students are regularly feeding back to teachers that specific structures and tailor-made revision tips are the most useful aid they can get, building confidence and enjoyment of maths for the forthcoming assessments and future mathematics courses, **Jones (2010)**.

Exam boards suggest that such practice tests will offer students the chance to see:

- Their test score
- Test analysis
- A test review showing model answers
- Skills map report showing performance vs. relevant GCSE specification topics

They are not practice exams, i.e. they are designed primarily to assess students' mastery of GCSE skills and highlight areas of strength and weakness in students' understanding to enable teachers to adapt schemes of work and revision accordingly, **Cumming (2010)**. The tests have the facility to block model answers, although when using such tests for teaching and learning, not letting students view and learn from mistakes, has limitations.

GCSEs delivered by onscreen testing offer a real benefit to students, and regularly receive very positive feedback from them, as they enhance exam delivery as well as students' enjoyment of the assessment process, **CIE (2009)**. In some international centres, such onscreen GSCEs on practical elements of a subject like Geography can also engage students in an area they would otherwise not be able to study (e.g. a practical test onscreen on rivers in a river-less country like Kuwait), **Wood (2009)**. The transition to summative e-assessment for onscreen GCSEs also makes such practice and familiarity of assessment useful for students. From June 2011, a number of exam boards are offering computer-based tests as an alternative to traditional paper-based assessments, and in some subjects / courses the computer-based test is the only option for certification (OCR GCSEs in Law, Environmental Science).

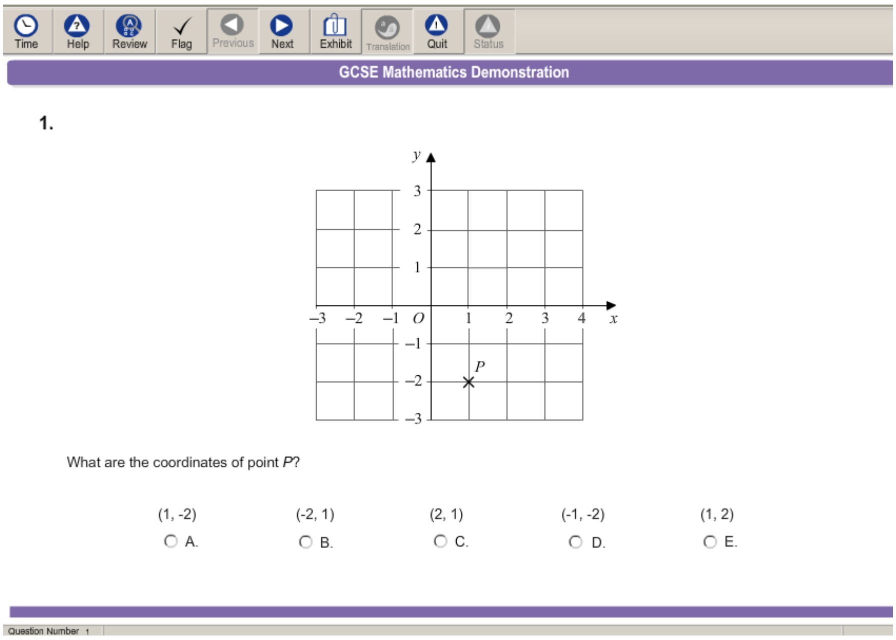


Figure 1.2: GCSE maths sample question for onscreen test, Edexcel (www.edexcel.com)

A full study of such onscreen GCSEs is warranted, especially the determination of the options for students to choose from, and is left to the discussion. At first glance above (in figure 1.2), it is clear to see that the five options are pedagogically the same, and that they don't give away any intuitive guesses for the candidate, which will increase their usability in the future. Our discussion throughout later chapters will centre heavily on the choice of options for the multiple-choice test.

The other area where some students will be exposed to multiple-choice testing is again via paper-based format, through the University of Oxford mathematics aptitude test (www.ox.ac.uk/maths) in figure 1.3 (below):

E. The cubic $$x^3 + ax + b$$ has both $(x-1)$ and $(x-2)$ as factors. Then (a) $a = -7$ and $b = 6$, (b) $a = -3$ and $b = 2$, (c) $a = 0$ and $b = -2$, (d) $a = 5$ and $b = 4$.	**H.** How many solutions does the equation $$2 = \sin x + \sin^2 x + \sin^3 x + \sin^4 x + \ldots$$ have in the range $0 \leqslant x < 2\pi$? (a) 0, (b) 1, (c) 2, (d) 3.

Figure 1.3: Sample questions from the University of Oxford mathematics aptitude test

While many students won't see these tests, they are another illustration of the power of multiple-choice testing to help differentiate a cohort of students, and coupled with the nature of accessible foundation level GCSE questions to aspiring University of Oxford mathematics undergraduates, show how diverse a span of abilities and mathematical experiences can successfully be summatively assessed using such question formats.

1.9 Diagnosis

The LTSN maths team published a guide (2002), on diagnostic testing for mathematics. It is interesting to remark how many university departments were using diagnostic tests then, and a similar study now would reveal just how many more were doing so. The striking point about the 2002 study was not that many departments still used a paper-based test, but how well versed in the pedagogy behind such tests they were.

At York University, **Todd (2002)** published a study (for electronics and physics students) on the correlation between A-level maths grades and subsequent performance, and revealed some striking results. Based on a fifty question paper-based multiple-choice test (that had not changed in format, style or substance over 15 years), they found the average mark (out of 60) dropped from around 40 (1986) to around 20 (2000), while A-level scores overall maintained a very similar level. They found the more marked decline amongst students who had scored a grade B at A-level maths, to the extent that the average grade B mathematician scored only marginally more on their test than they would have done so by guessing alone. Despite the fact the syllabi have changed slightly in content (if not really in style of questions) over this time, they found it impossible to explain the year-on-year decline in results.

By April 2001, despite the fact that huge numbers used diagnostic tests, only 18% of higher education institutions that responded to an LTSN survey used an on-line diagnostic test. Many had used paper-based tests for years, and a number didn't see the need to offer online tests. A similar survey of schools would have shown a similar or lower proportion using any formal notion of diagnostic tests, as didactic, teacher-led learning was still very much at the fore in the classroom.

At tertiary level, in this past decade we have seen the evolution from paper based tests to computer-based testing, **Robinson & Croft (2003)**. It has always been a learning objective to improve the mathematics education of students with a focus on preparation for tertiary transition. Such improvement, whether though drop-in surgeries, e-learning tasks, wikis, podcasts and the like are needed before students start undergraduate courses, yet often only undertaken once a student has started their courses. Such testing is not an end in itself, but only the beginning of a process of supported student learning, **Lawson (2003)**.

Technology has been key to the development of such testing programmes, and the development of a robust summative e-assessment framework has meant that spin-offs such as diagnostic testing at the start of a course, e-learning resources and online feedback have all become ideally suited to boosting individual learning outcomes, **Ross (2011**). Even the running of Open Education resources (OER), You Tube Edu, iTunes U has added to the rich number and variety of tools available to the university to appeal to learners of the new decade, **Rowlett (2010)**.

The need to tailor-make the diagnostic test is very much upon us. The next chapter aims to introduce the evidence from analysis of results files, with the twin aims of diagnosing basic errors in students' attempts at calculus questions and the development of sustainable (reusable) questions. This represents a partial solution to the problem of changing the nature of assessment, as getting students used to diagnostic testing has additional benefits: namely that they will be used to receiving feedback and also used to the e-assessment strategies employed in diagnostic testing, which will serve them well for the development of future summative e-assessment. **Greenhow et al. (2003)** pushed ahead with an e-diagnostic testing tool (Mathletics), with a user-friendly interface, and where questions are pooled together to make up repeatable tests (with marks stored online for comparison and teacher viewing). It would also offer direct feedback after each question, and suggested materials (or indeed retests) for students to work on in their own time.

1.10 Features of a diagnostic test

In designing a suitable test, we should use the following guiding principles: **Lawson (2002), Dalton (2006)**

- To inform staff of the overall level of competence in basic mathematical skills of the cohort they are to teach
- To determine whether a course of further study is right for the student concerned.
- To feedback / inform students of any gaps in the level of mathematical knowledge they will be assumed to have – so that they can take action to remedy the situation
- To be able to reuse the test for members of the same cohort and those in successive years, carefully refining the test to account for changes in learners entry profiles.

In follow up, schools and colleges should have in place many of the following to help students post diagnostic tests (without which there is little achieved, but to demoralise the weakest students):

- Maths surgeries
- Extra tutorials / classes
- Advice on which e-resources they can use
- Close monitoring / re-assessment
- Provision of paper-based resources / texts
- Advice on alternative courses / modules

1.11 Feedback

Nicol & McFarlane-Dick (2004), proposed the following seven principles of good feedback practice:

- Helps clarify what good performance is (goals, criteria, expected standards)
- Facilitates the development of reflection and self-assessment in learning
- Delivers high quality information to students about their learning
- Encourages teacher and peer dialogue around learning
- Encourages positive motivational beliefs and self-esteem
- Provides opportunities to close the gap between current and desired performance
- Provides information to teachers than can be used to help shape the learning

There is a discrepancy about when students can take full ownership of their learning, and some pre-GCSE age students may not be mature enough to cope with the demands of plugging the gaps themselves, and neither should they. However, there is growing anecdotal evidence from colleagues in secondary school maths teaching that students are keen to plug gaps as they approach and pass through GCSE and are keen to know which areas they need to work on first-hand. Even weak students at GCSE level know what they can and can’t understand.

Mathletics as written by **Greenhow et al. (1999-)** provides extensive feedback to the user on each multiple-choice question, and for each topic, based on Nicol & McFarlane-Dick's 7 principles for good feedback. The page is set up to provide an instant answer to a question, its full working (with associated general theory for the solution to a pedagogically congruent question), and the specific process / reason why the student went wrong. The feedback is designed also to help the student get to grips independently with the skills tested, so that they can repeat the same test very soon afterwards. There is a stark warning, however, in producing feedback for students as they work through tests, in that some will take too long looking through feedback without moving onto subsequent questions, **Gill & Greenhow (2006).**

1.12 Testing and measurement of efficacy of tests – a glossary of terms used.

Students make errors, and it is only by assessing these errors and analysing why and how often they make them that we can really make feedback and subsequent teaching resources useful to the individuals. This pedagogy must underpin our accurate writing of diagnostic materials. For each question written, there must be a number of learning weaknesses, and break-points from which students fail to accurately complete a question. By understanding and using these common errors, we can classify "**mal-rules**" as essentially rules or routes that a student as erroneously followed to generate a bad (mal) answer. Mal-rules are wrong paths producing errors which can be classified as mechanical, conceptual, procedural or by application, **Schechter (1994).**

We can also gauge students' ability to handle questions by measuring the levels of difficulty they face on it. Such a measurement, **Wood (1960), Crocker & Algina, (1986)** is defined by:

$$p_i = \frac{\text{Number of persons who answered it correctly}}{\text{Number of persons who answered it correctly and incorrectly}}$$

More simply, it is the mean mark obtained by a cohort divided by the total number of marks available any question, and is more commonly known as the "**facility value**" for that question. Should a question yield one mark, then the facility (F) will sit on a scale of 0 to 1, and this difficulty is really 1 - F. The nature of the question itself may well determine the facility value, as a poorly worded question may throw off genuine understanding, and a multiple-choice question with poorly chosen possible answers may equally allow a student to pick out the wrong choices without demonstrating any tangible mathematical skills. **McAlpine (2002)** proposed the following scale for facility value, to distinguish question facility:

$0 \leq F < 0.5 \rightarrow$ Hard

$0.5 \leq F < 0.7 \rightarrow$ Moderate difficulty

$0.7 \leq F \leq 1.0 \rightarrow$ Easy

Probably the most useful measure for a question designer is the "**discrimination index**"; in effect a measure of how well a question differentiates between members of a cohort. It will allow us to test the extent to which question responses discriminate between individuals who have a high score on the test and those that get a low score. The discrimination of a question is also a measurement on a sliding scale of -1 to 1, which shows how that particular question ranks compared to every other question in the test sat and it is obtained statistically as the Pearson product-moment correlation coefficient between the scores on the item and the scores of the total test, **Lomax (2000).**

It is worth pointing out that as the facility value approaches either 0 or 1, then the discrimination index will tend to zero quickly. An impossible question or far too easy question will not discriminate between candidates. **Massey (1995)'s** classification (figure 1.4) according to their discrimination will help guide the efficacy of questions in tests. While this table postures useful ranges of discrimination index for CAA in general, I firmly believe that it is applicable to mathematics assessments.

Discrimination index (D)	Quality	Recommendations
>0.39	Excellent	Retain
0.3-0.39	Good	Possibilities for improvement.
0.2-0.29	Mediocre	Need to check / review
0-0.19	Poor	Discard or review in depth
<0	Worst	Definitely discard.

Figure 1.4: Description of efficacy of the discrimination index, **Massey (1995)**

1.13 Nature of problem in transition

There are many other questions that need raising before we embark on a full thesis discussion of the role of e-assessment and e-learning. The advent of comprehensive databases in schools has considerably eased the ability of departments like the Centre for Evaluating & Monitoring (CEM) at the University of Durham (http://www.cemcentre.org/) to compile value-added analysis of a school's results year by year. This analysis is only as good as the objective tests at the start of a student's arrival at the school and the summative assessment they sit as they leave. Each are graded using paper-based tests, and this too would lend itself to technological advancement through the coming years, and will surely progress that way as summative assessment becomes computer aided.

Universities have been much quicker on the uptake of e-assessment, and it is clear that many of them have not only run diagnostic tools as part of their e-learning resources, but also successfully run end of semester testing for a number of years in many subjects. Such summative assessment has allowed a rich harvest of learning from e-resources and a major influx of e-resources developed in conjunction with those who write the summative assessments so that students can gain the maximum from these e-learning tools. Schools and colleges are still at the mercy of QCDA (and its predecessor QCA) and, while a number of them have adopted the initial pilot e-assessed GCSEs in maths, the wholesale delivery of e-assessment is a long way off still. Thus, the change from traditional assessment in schools to CAA in tertiary education will be a significant change for many students. This needs careful monitoring to make sure that students can be assessed over both assessment forms.

Chapter 2 - Methodology

2.1 E-assessment

My personal motivation for exploring e-assessment comes through contact with the University of California at San Diego, and their Department of Mathematics Testing and Placement (2001). They set out their aims for their diagnostic testing programme as:

- Setting tests for all entering undergraduates intending to begin calculus courses
- Analysing records of results to counsel students effectively
- Guiding students onto appropriate courses given their attainment levels
- Providing test materials to pre-college teachers to improve the mathematics education of their students with a focus on preparation for college.

As my understanding of diagnostic testing has developed, so I now see that it is not an end in itself, rather the beginning of a process of supporting student learning. In exploration of how scientists learn maths, I joined the IMA's conference on the mathematical education of engineers, **IMA (2003)**, and saw that in some universities, diagnostic testing was happening for over 1000 students per year, and testing was already becoming computer-based as opposed to the traditional paper-based assessments, **Robinson & Croft (2003)**. Moreover, my most significant discovery from this conference was how important the pedagogy of designing effective (online or paper-based) questions was in achieving learning objectives.

Technology has marched on significantly since 2003, and these enhancements are already underpinning many forms of summative as well as formative assessment in schools and universities. Computer-Aided Assessment (CAA) is now regularly embedded in the framework for learning in these institutions, as well as Computer-Aided Learning (CAL) both of which are more commonly known as e-assessment and e-learning. The pros of using e-assessment are many and far-reaching, **McKenna**

& Bull (2000), but the most important criteria for their successful implementation, **Paterson (2002)** are: Reliability, Validity, Usefulness, Fairness and Cost-effectiveness. In addition, I would add that they should be Sustainable (reusable), Interoperable, Instantaneous in marking and feedback and Relevant to future skills needed in life outside of the generic subject specific learning objectives.

Consequently CAA does need to be very carefully generated, as a non-repeatable e-assessment will not gain us much over its traditional paper-based counterpart. The ability to use a database and automatically generate hundreds or thousands of pedagogically identical questions at a mouse-click will allow us to use these tests for e-learning, and to offer tests to students in the same computer room such that they would gain nothing if they could see a neighbouring screen as it would have differing numbers in the question(s). Thus direct copying is eliminated, but other forms of cheating still exist such as aliasing, use of other software or resources.

Within online objective tests, the ability to obtain the best distracters (incorrect answers that can be presented to students as possible correct answers) and understand student mistakes is vital to gain an advantage over the paper-based test (which may have detailed teacher feedback), as this can allow instantaneous tailored feedback, with specific analysis of skill deficiencies and specific mistake identification. If these tools are in place, it will only help improve the efficacy of questions and the ability for such tests to discriminate between students in the future, **Earl, Land & Wise (2000).**

2.2 Test statistics and hypothesis testing

In the course of this research, it will become obvious that we will generate summary statistics from our various samples, and wish to compare them to see whether they are likely to come from the same parent population, or whether we detect a statistical difference between them, such that we can infer differences between the samples and offer unique recommendations for future analysis of the distinct samples. Given that the samples we take are independent, but not likely to have the same size, variance or likely to come from parent normal distributions, we can rule out using a t-test to compare the sample means. However, a **Mann-Whitney (1947)** test on the differences between the means of the samples can be carried out.

Thus a null hypothesis, H_0 would state: *there is no significant difference between performances on two sample sets*. as opposed to an alternative hypothesis, H_1: *there will be a significant difference between performance of the two sample sets*. We assume data to be taken uniformly from both sets.

This is a non-directional (2-tailed) hypothesis test, and as a first assumption, I will look to carry out the test at a 5% significance level, so that we have a reasonably good degree of confidence in our conclusions. We use a Mann-Whitney test, because it is a non-parametric test for assessing whether two independent samples of observations come from the same distribution. For example, suppose we want to test whether or not the samples of answers for the questions from set A are from the same distribution as those of set B. Then we can test whether the samples of answers for the questions from set B are from the same distribution as those from set A.

If our sample sizes are less than 20 for each sample, we won't have any justification for using a normal approximation to the distribution, but we can find (at a specified non-directional significance level, e.g. 5%) a test statistic, U, to be compared with tables of critical values for the Mann-Whitney U test: http://math.usask.ca/~laverty/S245/Tables/wmw.pdf

So, if U is above this critical value, we can reject H_0 in favour of H_1, that there is a statistical difference between the two samples. We can then adjust the significance level of the test downwards, e.g. to 0.01 (still non-directional), and compare again with the new critical values to see if we would draw the same conclusion from a more stringent test and confidence interval.

Table A5.07: Critical Values for the Wilcoxon/Mann-Whit

Nondirectional α=.05 (Directional α=.025)

	n_2																
n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45

Table 2.1: Critical values for the Mann-Whitney U test at the 5% non-directional level.

2.3 Correlation

Within our samples, we have other reasons for comparison, namely that of correlation. It is of great interest to us to see if a sample of results from one group or year is indicative of either similar or future success, so the ability to see how well the data from group 1 fits with that of group 2 will allow us to examine links between the two. There are two common mechanisms for calculating correlation, and both depend on whether we have a good hypothesis that the data is already linked. **Pearson (1917)**'s Product Moment Correlation Coefficient (PMCC) for the covariance of two sets of data relies on them to sit close to a linear regression line, **Galton (1888)**. Once we can make this assumption, then the PMCC is a simple calculation on Excel, that allows us to draw conclusions about the two sets of data in comparison. We can also use **Spearman (1904)**'s analysis on the correlation between two variables, and the assumption here is only that the two sets of data are rank ordered. In practice, both calculations often yield a result within 1% difference of each other, so allow an identical conclusion to be inferred from the comparison.

2.4 Use of technology in the thesis

Alongside the use of paper-based tests written on Microsoft Word, there are a number of diagnostic tests and analysis worksheets I have written using Microsoft Excel, because of its versatility to generate random numbers, produce test statistics, and to display tables, charts and graphs. On the accompanying CD is the "**Diagnostic test multi-use version.xls**", which is the pilot diagnostic test developed using Excel from chapter 5's analysis, and which demonstrates some of the pedagogy of writing objective test questions using the visual basic structure of Excel's worksheets.

The Edexcel exam board has also produced its own database, ResultsPlus, with dedicated centre login information for schools and colleges, and access to direct results for each student from any recent summative assessment completed from January 2008 to the present day.

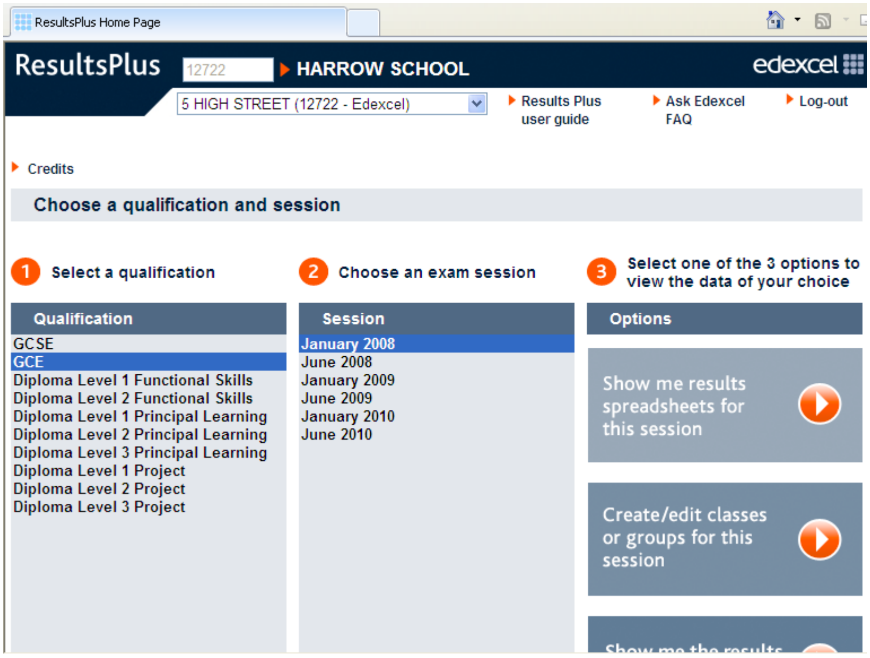


Figure 2.2: Edexcel's ResultsPlus website database selection page

Once inside the website database, I can download any exam paper report from each of the sessions listed above, and I receive a feedback worksheet. The feedback of results from the database is via a CSV file, which is easily exported to Excel for analysis, and once imported into Excel's worksheets, it is readily used to produce further test conclusions, summary tables and statistics relevant to the analysis in chapter 6. A full version of the feedback worksheets is also on the accompanying CD, and sample pages are described in section 6.1. Excel is also an invaluable tool in constructing the summary statistics used in completing the hypothesis tests in chapters 7 and 8 for the Mann-Whitney U test.

2.5 Samples used in the thesis

A sample is a finite part of a statistical population whose properties are studied to gain information about the whole, **Webster (1985)**. Research conclusions are only as good as the sample that we base those conclusions (or generalisations) on. Taking a sample of a parent population is by definition easier, less time-consuming, more cost effective, and simpler administratively. The samples used in this thesis come from three sources:

- 1st year Economics undergraduates at Brunel University, studying the EC1005 course (2008-2009 and 2009-2010)
- Year 11 and 12 mathematics students at Warwick School (2002-2003) studying a transition from GCSE to AS course in mathematics
- Year 11 and 12 mathematics students at Harrow School (2003-2007) studying a transition from GCSE to AS course in mathematics, and (2008-2010) studying an AS course (Core Mathematics 1).

Brunel University is a comprehensive university with approximately 14000 students, about 2000 of whom are overseas. It is therefore assumed that the university will be a good representative of UK higher education institutes and that the sample will give a true indication of the majority of undergraduates studying maths-related courses where mathematics diagnosis may be required at an early stage in their first year, **Baruah (2007)**. Warwick and Harrow schools are independent secondary schools, both with male only cohorts. The nature of selection from these institutions is primarily for ease of access of a large quantity of sample data, and despite this yielding less certain generalisations of a parent population, we can compare with data generated through a random sample of schools, colleges and other further education centres, as in section 3.3.

Chapter 3: Determination of student errors in calculus questions

3.1 Core Maths 1 scripts from Results Plus

Using 374 “ResultsPlus” answer files from Edexcel, the associated Core Maths 1 (C1) exam papers and their mark schemes, I have put together an analysis of 9 differentiation questions sat by candidates at one exam centre and will pick out some likely errors from our data (all of the results analysis files are available to view on the CD). More than half of the students scored 100% on the topic questions, and all bar 100 papers had 80% or higher scores on the differentiation topics alone. It is these 100 papers and their associated errors that I will pursue in depth. Differentiation, in fact, remains a very well understood topic, across the years as this table compares the facility of differentiation questions, with the rank facility of the 12 topics tested on each exam paper:

	Jun-07	Jan-08	Jun-08	Jan-09	Jun-09	Jan-10		Overall
Facility (out of 1)	0.95	0.90	0.90	0.93	0.81	0.83		0.89
Rank (out of 12 topics)	10	12	6	8	1	3		10
Marks available	13	14	15	13	18	11		14

Table 3.1: Facility of differentiation questions in C1 exam scripts, Jun 07 to Jan 10

The examination in June 2009 caused greater issues, with a larger than average number of marks available (18 out of 75 total, verses a long term average of 14), and all possible syllabus sub-topics were examined with somewhat tricky, fiddly questions.

When we break down the individual marks on each part question into method marks (M1), or answer marks (A1), we see can see more detailed patterns emerge. The table below shows the marks awarded to each part question per year – I have focussed on June 08 to Jan 2010 and on the questions that require polynomial differentiation (including fractional and negative powers), thus leaving aside analysis on questions requiring the skills of finding tangents and normals to polynomials at give points. It is for low facility values on these individual marks where I can focus my attention to the feedback and exam scripts to see where mistakes have been made.

Jan-10	Q1 M1	Q1 A1	Q1 A1	Q6a) M1	Q6a) A1	Q6a) M1	Q6a) A1
Correct	35	34	24	33	29	32	28
Mistake	0	1	11	2	6	3	7

Jun-09	Q3a M1	Q3a A1	Q3a A1	Q9b) M1	Q9b) A1	Q9ba) A1
Correct	81	78	77	78	72	75
Mistake	0	3	4	3	9	6

Jan-09	Q6b) M1	Q6b) A1	Q6b) A1	Q6b) A1	Q11a) M1	Q11a) A1
Correct	39	38	35	36	38	37
Mistake	0	1	4	3	1	2

Jun-08	Q4a) M1	Q4a) A1	Q9a) M1	Q9a) A1
Correct	107	107	107	104
Mistake	0	0	0	3

Jan-08	Q5b) M1	Q5b) B1	Q5b) A1	Q5b) A1
Correct	12	11	11	11
Mistake	0	1	1	1

Figure 3.2: Analysis by question mark for differentiation of polynomials.

3.2 An analysis of student errors through examiners' reports

The following analysis is based on the individual scores from the ResultsPlus files, and insight from Examiners' Reports, GCE Mathematics (8371/8374/9371/9374), Edexcel (2008-2011). www.edexcel.com

June 08: Q9a) Differentiate: $y = kx^3 - x^2 + x - 5.$

Of the 107 attempts, only 3 failed to score both marks, and given that they all scored the method mark (available for finding 1 or more correct terms), then the presumed mistakes are almost certainly mechanical, as one new power had to be correct, which indicates knowledge of the techniques. The report cited errors with the “k” term, and errors of the form: kx^2 or $3k^2x^2$ or $3kx^3$ were all observed.

January 09: Q11a) (By first differentiating the function), find the gradient of the tangent at (2, -3): $y = 9 - 4x - \frac{8}{x}$.

Of the 38 attempts, only 2 failed to get full marks, and one of those gained the method mark. It is possible that the full question (which required the tangent gradient and didn’t explicitly ask for the derivative), may have not helped those who didn’t have the knowledge of the technique for finding the tangent, yet could have differentiated the function above if asked specifically for it. It is more likely that this happened to one candidate, while the other (1 mark out of 2) candidate made a generic mistake with the final term. The popular mistakes were noted in the differentiation of the final term with: $\frac{8}{x^2}$ or $\frac{8}{x}$ seen often.

Jan 09 Q6b) Differentiate: $y = 5x^4 - 3 + 2x^{\frac{3}{2}} - x$

Of the 39 entries, 34 successfully completed all four marks. The other 5 candidates all scored at least two of the four marks (including the first method mark and one answer mark), so clearly lost out with mechanical errors not complete conceptual / knowledge based errors. The major challenge in this question was to differentiate the fractional power term correctly, and I presume that those five who lost one or other mark would have lost them at this point. The common mistakes seen for differentiating $2x^{\frac{3}{2}}$ were: $3x^{\frac{3}{2}}$, $2x^{\frac{1}{2}}$, $\frac{4}{3}x^{\frac{1}{2}}$.

June 09 Q9b) Differentiate: $f(x) = 9x^{-\frac{1}{2}} + 16x^{\frac{1}{2}} - 24$

This was the hardest part question of that paper, and 3 candidates of the 81 left it blank (scoring 0 out of 3), or made a major error early on in the question that they didn't score the method mark (gained for dropping the power of one term by 1). Such errors are more than likely mechanical, as all 3 of them scored the full 3 marks on the harder derivation previously to get to the stage above, demonstrating excellent algebraic skills and competency with indices. The question offered one follow-through mark for those who had derived the wrong expression above, but only one of those 7 who failed to score full marks would have needed it. Those 7 scored a method mark, but they lost one or both answer marks for the two main derivatives. Each one needed simplifying, but because the simplifying (16 x ½ or 9 x ½) isn't as difficult a task compared to the differentiation in my view, then it is likely that their errors were mechanical mistakes with indices.

Commonly seen errors for differentiating $9x^{-\frac{1}{2}}$ were: $4.5x^{\frac{1}{2}}$, $4.5x^{-\frac{1}{2}}$, $4.5x^{-\frac{3}{2}}$, $-4.5x^{-\frac{1}{2}}$, $-4.5x^{\frac{1}{2}}$ and for differentiating $16x^{\frac{1}{2}}$ were: $32x^{\frac{1}{2}}$, $32x^{-\frac{1}{2}}$, $8x^{\frac{1}{2}}$

June 09 Q 3a) Differentiate: $y = 2x^3 + \frac{3}{x^2}$

The 81 candidates who sat this question made short work of it with the exception that 7 of them dropped a mark at one of the answer stages. Given that the first method mark (M1) was awarded to all 81 for showing one or more correct power, then we can reasonably clearly identify the candidates' mistakes on each term through the examiners reports for those that dropped this mark. In fact this question would lend itself to a numerical input type question, whereby the coefficient and power offered by the student would help us tell immediately what sort of mistake was made (given a set of 7 or 8 highly likely mistakes as above). For the candidates who got the derivative of the cubic term wrong: $2x^2$, $\frac{2}{3}x^3$ or $6x^3$ were spotted by examiners. Similarly, those that got the negative power wrong were most likely to have offered: $6x$, $6x^2$, $-6x^3$, $-6x^{-2}$ or $6x^{-3}$ instead.

Jan 10 Q1) Differentiate: $y = x^4 + x^{\frac{1}{3}} + 3$

At first glances, this was a very innocuous looking question, yet yielded a very poor success rate for the 3rd answer mark. For the 35 entries, all scored the method mark (for one correct power), and one answer mark for the "0" or "$4x^3$" but 11 of the 35 failed to get the answer mark for the middle term. Clearly the subtraction of 1 from 1/3 is harder than it looks, and the mistake was made here. The examiners were quick to point out this major error universally, and the major mistakes were: $\frac{1}{3}x^{\frac{1}{3}}$, $\frac{1}{3}x^{-\frac{1}{3}}$, $\frac{1}{3}x^{\frac{2}{3}}$, or $x^{-\frac{2}{3}}$

3.3 Evidence from different centres in 2002

During 2002, I spent time employed by Edexcel to mark old-style Core maths exam papers. The process involved the marking of around 400 papers entitled: Pure Maths 1 (P1) in January 2002 and Pure Maths 2 (P2) June 2002. From standardisation meetings, examiner team meetings and notes taken during the marking process, I have sampled two particular differentiation questions that showed up similar errors nearly a decade previously from a cohort of students drawn from across the nation. This has the advantage of being a random sample of student exam scripts, as the centres are allocated randomly to examiners. It won't show how errors have changed, as the style of questions were different (given a change in syllabus), and the emphasis and marks available were weighted differently as well.

P1 (Jan 2002): Differentiate: $y = x^3 - 5x^2 + 5x + 2$

One method mark was awarded just for one term where the power had been reduced by 1. Such rewards can prompt future mistakes, i.e. differentiating x^3 and yielding x^2. Similarly mistakes on this question were less common, but equally revealing, i.e. the "+2" often remained, and the 5x often stayed as 5x or disappeared. The answer mark for the question was for the complete answer only, so again didn't allow any slips.

P2 June 2002: Differentiate: $y = 10 + \ln(3x) – 0.5\ e^x$

Again this yielded very many good solutions, but high-lighted the mistakes of generally weaker candidates (a good discrimination index here) for differentiating the middle term. Most errors involving the first and last terms were to leave the 10 in (note the similar mistakes observed in January 2002), or to change the sign in-front of the 0.5 e^x term. The middle term should yield 1/x, but very commonly I saw: 1/(3x), 3/x or (1/3)x:

$y' = 1/(3x) – 0.5\ e^x$
$y' = 3/x – 0.5\ e^x$
$y' = 10 + 1/(3x) – 0.5\ e^x$
$y' = 10 + 3/x$ + or - $0.5\ e^x$

3.4 Core maths examiners ideas of mal-rules:

As well as reading the associated examiners' reports for each modular A-level exam, I have also attended a number of teacher training days run by Edexcel, for which examiners' will systematically dissect previous session exam papers and offer specific marking advice, summaries of generic errors across the country and outline their vision for assessing certain specific skill sets. It is from these reports and feedback meetings, I summarise some additional feedback on differentiation problems in Core Maths 1 from June 2009 and June 2008 sessions and IGCSE Maths sessions from June 2009 and June 2010.

- Differentiating expressions can sometimes yield an integrand (and/ or +C in the answer too)
- Differentiating $x^{\frac{1}{3}}$, yields: $\frac{1}{3}x^{-\frac{1}{3}}$
- Differentiating $y = \frac{(x+3)(x-8)}{x}$ proves hard, as sometimes the numerator and denominator are differentiated separately, or the final answer is

multiplied by x or 1/x if not simplified first. For some the multiplication to give $x^2 - 24$ proved too tempting. This again cropped up on an $(x^2+3)^2$ expansion.

- Differentiating $\frac{3}{x^2}$ gave the following wrong answers: $-6x^{-1}$ or $+6x^{-3}$
- Multiplying surds: $-4\sqrt{x} \times -4\sqrt{x}$ was often given with answers such as $-4\sqrt{x}$ or $\pm 16\sqrt{x}$ or $\pm 16x^{\frac{1}{4}}$. Dividing by $\sqrt{x}$ was equally hard, and some thought that $\frac{x}{\sqrt{x}} = 1$.
- When differentiating $\frac{2x^2 - x^{\frac{3}{2}}}{\sqrt{x}}$, candidates found the simplification hard, and some multiplied instead of divided, giving $2x^2 \div \sqrt{x} = 2x^{\frac{5}{2}}$, and sometimes $\sqrt{x}$ was interpreted as x^{-1}. Some candidates divided only one term by the denominator too.
- Differentiating simple polynomial expressions such as $f(x) = 3x + x^3$ were mostly good for the first term, but a few got the 2nd term wrong (values such as 2, ½ or 1/3 were seen).
- Harder expressions such as $y = kx^3 - x^2 + x - 5$ proved harder to differentiate. Some weaker candidates were confused by the k and answers such as $2kx^2$ or $3k^2$ were seen.
- Differentiation questions reward method marks for reducing the power, so some candidates have forgotten to multiply the power to the term in-front of the x.
- The misreading of questions continues, and some still confuse: $5x^{\frac{1}{3}}$ with $\sqrt[3]{5x}$, and confusing $(5x)^{\frac{1}{3}}$ with $5x^{\frac{1}{3}}$, or $4\sqrt{x}$ with $x^{\frac{1}{4}}$
- Beware a negative index meaning a negative result – too many students have spent years wrongly assuming that negative indices will automatically yield a negative result in whatever context.
- Confusing integration with differentiation – all too commonly there are students who will differentiate when given a question including an integral

sign, and equally when they see f'(x) will automatically differentiate without reading the question that may well ask them for f(x).

- Use of f'(x) – again this causes students confusion with the inverse function, some new function or is a piece of revision that slips the mind.
- Gradients of polynomials - students rarely know why the gradient of y = mx + c is "m", and equally have often confused the first coefficient of an expression with the gradient automatically, i.e. for y =ax^2 + bx + c, some have thought (wrongly) that "a" was the gradient.

3.5: IGCSE results plus analysis of student errors

Edexcel's IGCSE papers provide another source of data for assessing student errors with basic calculus. The syllabus for the IGCSE has a good overlap with that of the Core Maths 1 paper (June 2007 paper 3 had 58 out of 100 marks specifically on topics or skills that overlapped). Specifically, it expected candidates to find the 1st derivative for polynomial functions, including those with negative indices (but not fractional), and also the extension to finding turning points of basic functions. The same ResultsPlus data is available for candidates, but only by part question, i.e. yields a numeric score out of 4. What we can deduce from this is more to do with the actual question and its relative difficulty than the actual mal-rules.

June 2007 (paper 3H) Q17) : $y = x^2 + \frac{16}{x}$. The curve has one turning point.

Find $\frac{dy}{dx}$ and use your answer to find the coordinates of this turning point

Looking at the results of this question, it has a facility value of 0.600 for the whole cohort of 155 students, and when compared to the final scores (out of 58 marks assigned as C1 skill overlap), a discrimination index of 0.6283. Thus it is a good discriminator for those who wish to pursue maths further to GCE level and, given the relatively high average scores (facility of 0.84) per student for the whole exam paper, is considerably more difficult than other areas (ranked 1 out of 10). In fact:

	Quadrat	Indices	Shape	Formula	Polynom	Percent	Pythag	Inequal Lines	Simult-aneous	dy/dx
Marks	**14**	**9**	**8**	**20**	**8**	**7**	**5**	**4**	**6**	**4**
Facility	0.88	0.76	0.76	0.90	0.85	0.88	0.88	0.85	0.74	0.60
Difficulty ranking	7	3	4	10	6	8	9	5	2	1

Table 3.3: Topic analysis for 2007 paper 3H IGCSE for overlap C1 topics.

It is this that prompts me to look very carefully through the mark scheme for the question to work out where students are likely to have gone wrong: (table 3.4 below)

Marks (/4)	0	1	2	3	4
Frequency	9	13	76	17	37

Those who scored 0 marks clearly left the question out or failed to understand even the nature of differentiating, as writing down “2x” would have scored a mark (one mark available for reducing the power of any term by 1). Their errors are almost certainly a lack of knowledge of the whole topic. The group who scored 1 mark most likely scored it on the first answer mark (B1) for writing “2x”, so have shown some knowledge of basic polynomial differentiation. Their mistakes are likely to be more complex and harder to fathom for the second term and subsequent likelihood of blank answers for the turning point (again a lack of knowledge).

The majority of students scored 2 marks. If they were for the correct differentiation (scoring two answer marks, “B1 B1”, and nothing for finding the turning point “Method 0 Answer 0”), then their error on the 2nd requirement was that they had no idea how to start finding the turning point (or perhaps what it was). Thus the error was one of unfamiliar terminology or lack of knowledge rather than mechanical. It is possible that there were a number of students who scored B1 B0 M1 A0, so had knowledge of how to differentiate and achieved the correct answer with one term, and then equally had knowledge of how to start finding a turning point but failed to complete either exercise accurately afterwards, thus scoring the method mark alongside one answer mark. I am most interested in seeing their scripts, as it is highly likely that they wrote down the derivative as $2x \pm something$ and that something would yield further clues to the mechanical errors in differentiating a negative power.

Those who scored 3 out of 4 must have scored B1B1M1A0, i.e. only made the mistake at the very end. Clearly to jump from $2x \pm \frac{16}{x^2} = 0$ which has scored the method M1 mark to the final answer is a serious amount of work for one mark, and there are many errors to befall a student between there and the final coordinates. Fortunately there were a good number who scored all 4 marks, and thus demonstrated a very sound understanding of a number of key skills in one go.

June 2009 4H: Differentiate: $y = x^2 + 3x$, and then find the gradient when y = -4, and find the turning point.

This question had a similar facility value as two years previous (0.61), which was borne out both nationally and within our centre's students. Basic procedures followed to obtain 2 correct terms of 2x and 3, which were awarded the two marks, even if combined to make 6x or 5x incorrectly afterwards. This picks up the first likely error – that of combining unlike terms. Viewing sample scripts also showed up how x^2 often got differentiated to x rather than 2x as well. The 3x term also remained as 3x or disappeared on a number of occasions. The advice was clear - getting this wrong (even with ax + b left) can lead to 2 out of 5 overall.

In the second part: find gradient when y = -4, follow through only awarded when candidates substituted -4 into "ax + b" from (a), and common mistakes were seen when some substituted "-4" into the original equation. There were also common mistakes on substituting numbers in: 2(-4) + 3 = 11, or 2 (-4) + 3 = -11, were seen often. While these parts were well answered, finding the coordinates of a minimum point proved harder, and many blank responses indicated a lack of knowledge. Marks were only awarded that showed valid algebraic working, typically 2x + 3 = 0 (M1).

Both x and y values needed for the accuracy marks (A2, 1 each) – the advice was to regularly advise candidates to find BOTH answers (like simultaneous equations). Sometimes the use of $-b/2a$ would be seen for the turning points (and accepted, when evaluated correctly).

When looking closely at IGCSE scores on C1 overlap topics, we find that indices scored only 76% average (9 marks of the 58) and ranked 3rd most difficult topic, with Simultaneous equations ranked 2nd. The basics within the GCSE and IGCSE courses underpin the skill development through C1 and associated courses, and it is clear that candidates are unlikely to fall down on knowledge of technical aspects of a calculus question, but more likely to fall down on the mechanical side with poor manipulation of indices. Examiners have always pointed out that students are still struggling on calculus questions involving negative and fractional indices, and specific questions written to diagnose and help correct these weaknesses are ever more needed as one skill underpins another.

Chapter 4: Generating a taxonomy of errors and mal-rules

4.1: Learning styles

Learning mathematics is about understanding and applying the rules than make up the language of mathematics. Being a real mathematician is about making connections which are then demonstrated in the language of mathematics we produce and are able to understand, **Harte (2010)**. Defining a useful taxonomy for student errors, has also long been a goal of educators, as to second guess where these likely errors will occur will not only educate the educator but also guide the learner towards understanding of basic skills, application of increased knowledge and creativity in the subject. After all, many teachers use intuitive taxonomies to categorise student error – a method more often reflective of personal idiosyncrasies than actual mistake significance, **Anson (2000).**

Understanding learning outcomes and what learning mathematics actually means, is key to understanding how students make these mistakes. **Bloom (1956)** postulated his model for taxonomy of learning outcomes via three domains. Using the first (cognitive domain) of learning, he suggested that one cannot effectively — or ought not try to — address higher levels until those below them have been covered (it is thus effectively serial in structure). Thus we start by building up knowledge, then comprehension and follow this by comprehension. The upper three levels of the pyramid are often thought of as irrelevant for most secondary school students, **Davies (2010)**.

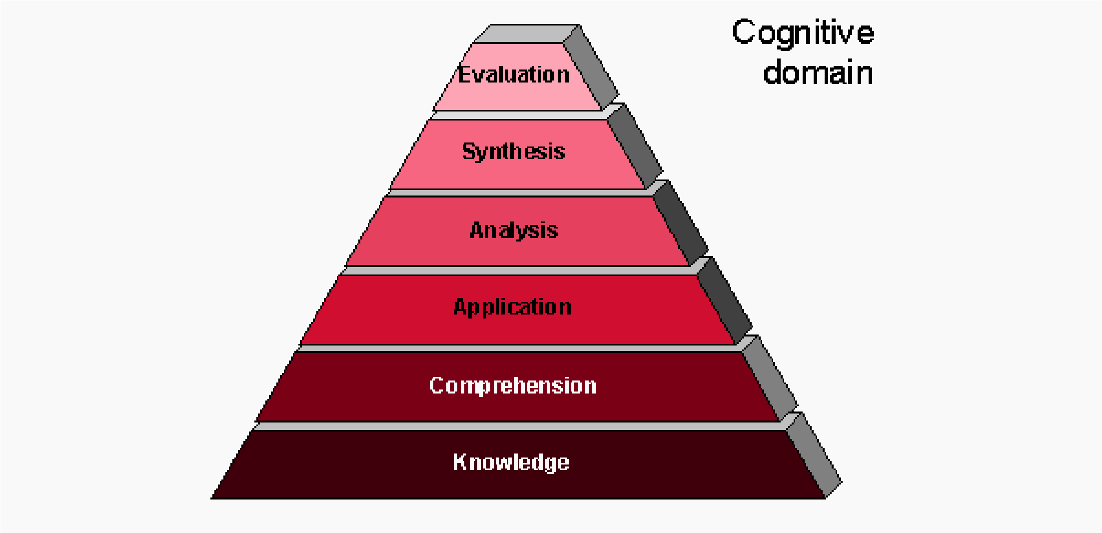


Figure 4.1: **Bloom (1956)**: Cognitive domain for learning objectives.

This taxonomy is a generalised set of categories, and mathematics teachers will often concentrate their teaching at the "knowledge" level, as a good mathematician needs to have a head full of techniques and formulae. In a similar way to learning a language, a good linguist needs to have a head full of vocabulary, verb conjugations, noun declensions etc, as without this mass of knowledge and the facility for instant recall, a linguist will not get very far, **Davies (2010)**.

Many secondary schools who use a taxonomy to guide student learning (or more often to satisfy the Ofsted inspectorate) have updated **Bloom (1956)** to read as follows:

• Know (Remember)	• Presentation	• Remember
• Understand (Concepts)	• Practice	• Understand
• Be able to (Apply)	• Production	• Apply

Table 4.2: Some taxonomies used in secondary schools to define learning objectives

Marzano (2000) updated Bloom's educational objectives into his "new Taxonomy" of three systems and the knowledge domain, mainly in response to syllabus guidelines-based instruction in table 4.3 below. Using his cognitive system, he defines four sub-classes to his system, much in the same way as Bloom does, using Knowledge, Comprehension, Analysis and Knowledge Utilisation (Application) as his headers. The system can prove useful in fleshing out Bloom's taxonomy, but still needs tailoring to suit the mathematical purpose to which I am about to use it.

Cognitive System			
Knowledge Retrieval	Comprehension	Analysis	Knowledge Utilisation
Recall Execution	Synthesis Representation	Matching Classifying Error Analysis Generalising Specifying	Decision Making Problem Solving Experimental Inquiry Investigation
Knowledge Domain			
Information	Mental Procedures	Physical Procedures	

Table 4.3: Marzano (2000) – a new taxonomy for learning objectives.

Of those who have attempted to redefine these taxonomies for mathematical learning, **Hatt & Baruah (2005**) put forward a version of Bloom's taxonomy based on: Remember, Understand and Apply. They used the pyramid idea from Bloom and defined four major classes of errors:

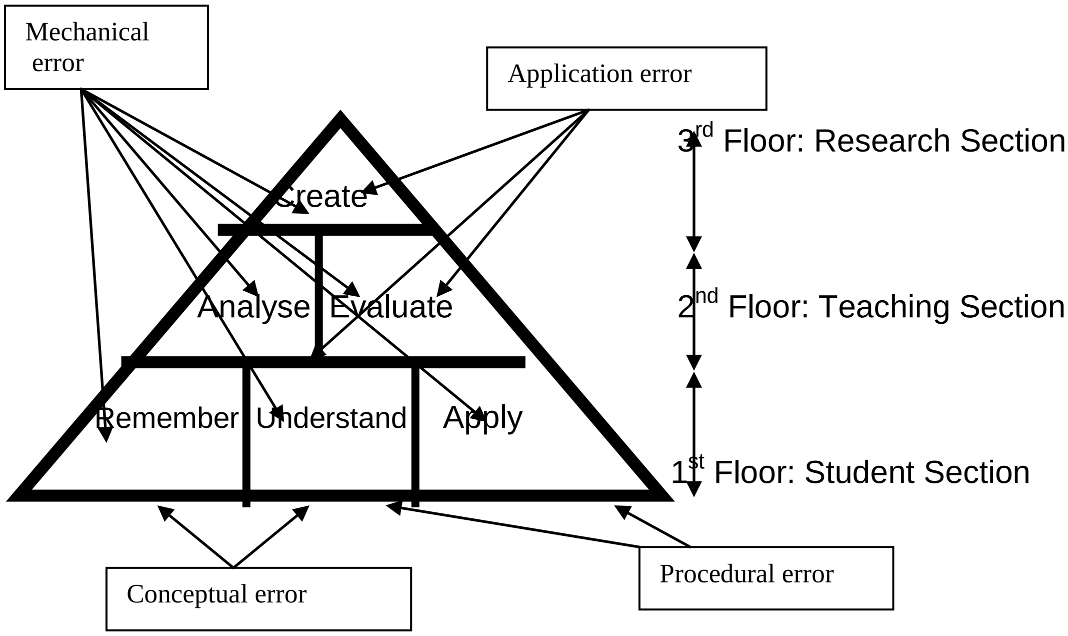


Figure 4.4: Error identification in various learning levels, **Hatt & Baruah (2005)**

4.2: Classification of students' mistakes

The issue surrounding this work is that the correct identification and classification of mistakes can be an arbitrary process, unless we use copious evidence of student work to be able to see exactly how they went wrong in their solutions. To build up a taxonomy from a theoretical basis is always going to be inaccurate, i.e. when mal-rules are based on learning misconceptions not different previous mistakes, as we can be guilty of subjectively leading question design and learning assessment to fit out own taxonomy, **Baruah (2007).** I feel most sure of what mistakes students make by seeking out their own work – hence my use of an evidence based taxonomy.

Orton (1983) also classified the students' errors in three broad classes as: structural, executive and arbitrary. **Structural** errors arose from failure to grasp some principle essential to the solution or appreciate relationships involved. **Executive** errors involved failure to carry out manipulations accurately or fully, even though the key principles or techniques involved may have been understood. **Arbitrary** errors were said to be those in which the student behaved arbitrarily and couldn't cope with the problem at all. He, like many others, showed that the understanding of a single concept or skill might be comprised of one or more different classes of error.

4.3: The taxonomy of SOLO

The taxonomy which I believe fits the mathematical world best is that derived by **Biggs & Collis (1982)** – SOLO (Structure of Observed Learning Outcomes), which divides understanding into five levels. The beauty of this structure is that questions can be tagged easily according to the level they aim to assess, and the mistakes from the learner are equally easily categorised according to achievement or lack of achievement of these levels.

- **Pre-structural**. An incorrect process is used in a simplistic way which may lead to an irrelevant conclusion. Or the student may even fail to engage in the problem, so there is no outcome at all.

- **Uni-structural**. A single process or concept is applied to at least one data Item (i.e. a learner has grasped one element of a concept). A conclusion is drawn, but unless the single process together with the selected data suffice for the correct solution of the problem, the conclusion will be invalid.

- **Multi-structural**. A number of processes or concepts are used on one or more data items, but with no synthesis of information or intermediate conclusions. This lack of synthesis may be acceptable in the case of a straightforward question, or may indicate cognitive performance below that required for successful solution of the larger problem, i.e. I have a pile of bricks but can't yet see how they fit together to build the house.

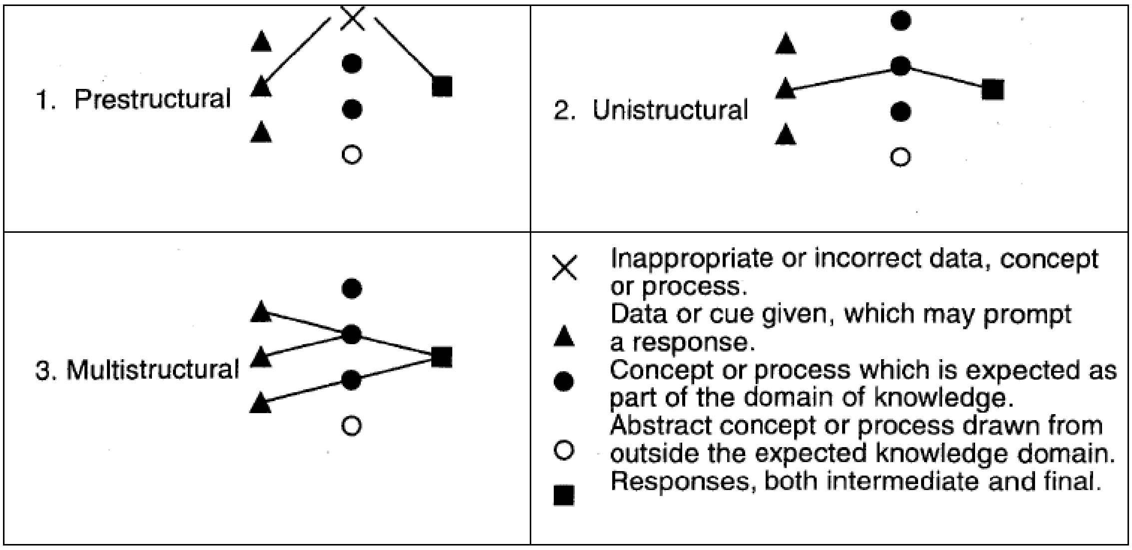


Table 4.5: Response structure at different levels of the SOLO model
Biggs & Collis, (1982), **Watson, (1994)**).

- **Relational**. A relational response is characterised by the synthesis of information, processes, and intermediate results. In order to achieve the conclusion, concepts are applied to some of the data, giving interim results which are then related to other data and/or processes (i.e. I can now see how to build the house from the bricks I have been given).

- **Extended abstract**. Extended abstract responses are structurally similar to relational responses, but here data and/or concepts and processes (more usually the latter two) are drawn from outside the domain of knowledge and experience that is assumed in the question.

Typically, how we can use SOLO is that teachers, in setting the problems, will provide the students with data in the problem (explicit cues), expecting the students to use concepts taught in the unit of work which may not be specifically identified in the question (implicit cues). Such implicit and explicit cues will be said to be in the student's domain of knowledge, **Chick (1998).** To exemplify the nature of this taxonomy, consider the range of questions depicted in the analysis in chapter 3. The analysis above can be summarised succinctly in the following table (4.6), whereby I have selected 12 commonly used terms in polynomial differentiation questions, and attached 7 common error types as evidenced through chapter 3 to each required derivative (the questions are given in the left-hand column). The error types are described below the table, and each error was evidenced through student work.

d/dx of	Error 1	Error 2	Error 3	Error 4	Error 5	Error 6	Error 7
$\mathbf{2}$	2	$2x$					
$\mathbf{5x}$	$5x$	0					
$\mathbf{2x^3}$	$6x^3$	$2x^2$	$2x^2$	$(½)\ x^2$	x^2	$(2/3)\ x^2$	
$\mathbf{kx^3}$	$3kx^3$	$3k^2$	kx^2	$2kx^2$	$(k/2)\ x^2$	$(k/3)\ x^2$	
$\mathbf{3x^{-2}}$	$-6x^{-2}$	$6x^{-3}$	$3x^{-3}$	$9x^{-3}$	$-x^{-3}$	$-(3/2)x^{-3}$	$-6x^{-1}$
$\mathbf{2x^{3/2}}$	$3x^{3/2}$		$2x^{½}$	$x^{½}$	$4x^{½}$	$(4/3)x^{½}$	
$\mathbf{-8/x}$	$8/x$		$-8/x^2$	$16/x^2$	$-4/x^2$	$\mathbf{8/x^2}$	
$\mathbf{9x^{-½}}$	$-4.5x^{-½}$	$-4.5x^{½}$	$9x^{-3/2}$	$(27/3)\ x^{-3/2}$	$-6x^{-3/2}$	$-18x^{-3/2}$	
$\mathbf{16x^{½}}$	$8\ x^{½}$		$16x^{-½}$	$-16x^{-½}$	$-32x^{-½}$	$32x^{-½}$	
$\mathbf{x^{1/3}}$	$(1/3)x^{1/3}$	$(1/3)x^{-2/3}$	$x^{-2/3}$	$-(2/3)x^{-2/3}$	$-(3/2)x^{1/3}$	$3x^{1/3}$	$\frac{1}{3}x^{-\frac{1}{3}}$
$\mathbf{ln(3x)}$	$3/x$	$(1/3)x$	$1/(3x)$				

Table 4.6: An error analysis table for differentiation of polynomial functions

- Error 1: No reduction of the power, but correct multiplication of the old coefficient by the old power
- Error 2: Confusion with definitions (e.g. 5x having the same derivative as 5, or 2 having the integrand of x^2 – recognition of something similar being the result, or sign error).
- Error 3: Correct reduction of power, but no multiplication of the old coefficient by the new power. The new coefficient is the same as the old coefficient.
- Error 4: Correct reduction of power, but multiplication of the old coefficient by the new power.
- Error 5: Correct reduction of power, but division of the old coefficient by the new power.
- Error 6: Correct reduction of power, but division of the old coefficient by the old power. Notice how this “error” leads to the correct result in the differentiation of ***-8/x***
- Error 7: Additional mistakes pointed out by examiners that look intuitively correct, but involve raising the power or wrongly subtracting 1 from a fraction.

Each of these question types falls under the remit of needing a multi-structural response, as there are responses that show uni-structural levels but are not completely correct. Indeed, many similar questions (facility value) will be multi-structural problems, because they require the student to demonstrate a number of processes on one or more data items. They have to complete the differentiation of a polynomial function, which requires implicit skills from their domain of knowledge (recognise the need to differentiate, re-write the question if needed with fractional or negative power prominently displayed, reduce the power of the x term by one, multiply the coefficient by the original power, simplify the final expression).

Error 3, 4, 5, 6, 7	Uni-structural response	A single process or concept has been grasped (i.e. power reduction or manipulation of the coefficient) but the conclusion drawn will be invalid.
Error 2	Pre-structural response	An incorrect process is used in a simplistic way leading to the wrong solution or no solution.
Error 1	Pre-structural response	An incorrect process is used in a simplistic way leading to the wrong solution or no solution.

Table 4.7: Summary of SOLO tags to errors on polynomial differentiation questions

Once we recognise that some answers are pre-structural in level (to a multi-structural level question), then we can flag up warnings that the students need to undergo some further testing to confirm these results or go straight back to remedial work on this skill development. While there is no concept of a pre-structural question, only pre-structural error, it does make sense to include the notion of uni-structural problems, and worth noting that setting too many uni-structural problems will leave the students often unable to solve similar questions in different representations, **Robin & Rider (2010)**. If we set up assessments for students based on these levels, and could analyse the responses accordingly, then we can assign an overall level to students based on how well they achieved learning outcomes for this skill, **Watson, Chick & Collis (1988)**.

4.4: The use of taxonomy in deriving mal-rules

Priem (2010) also focuses on mistakes from e-learning, and notes how little work has been done to answer the question posed at the start of the thesis: whether students make the same mistakes on paper-based tests as online. He proposes a three-fold taxonomy to error classification for e-learning tools: Learning mistakes, technology mistakes (e.g. closing browsers, submitting solutions prematurely) and general mistakes. Re-categorised each area as with SOLO; pre-structural mistakes (learning mistakes and general mistakes) or uni-structural mistakes (technology mistakes), which leaves us really with only two error levels in this model. However he still comes to the same conclusions, that examining error types will help students plan strategies for error reduction and encourage appreciation of valuable learning errors.

Schecter (1994), Orton (1983), and Greenhow (1996), all show through their independent research that the same errors are found in higher dependency skills, i.e. an error in basic differentiation shows up again as same error for a question involving the solution of an Ordinary Differential Equation (e.g. using a Laplace transform). In fact, **Baruah (2007)** also demonstrated through her work that errors in differentiation questions on polynomial functions led directly to similar errors in differentiation on functions that required use of the product, quotient and chain rules.

The point is shown in table 4.8 (below), where I use a simple content course dictionary for differentiation in the core maths 1 (C1) module. It is clear to see the interdependency of many GCSE skills and topics even before differentiation is first explored with students, and as such, many students will make several of these uni-structural errors on differentiation of polynomial functions, because their knowledge, understanding and ability to apply rules about indices, surds, straight lines etc is patchy at best.

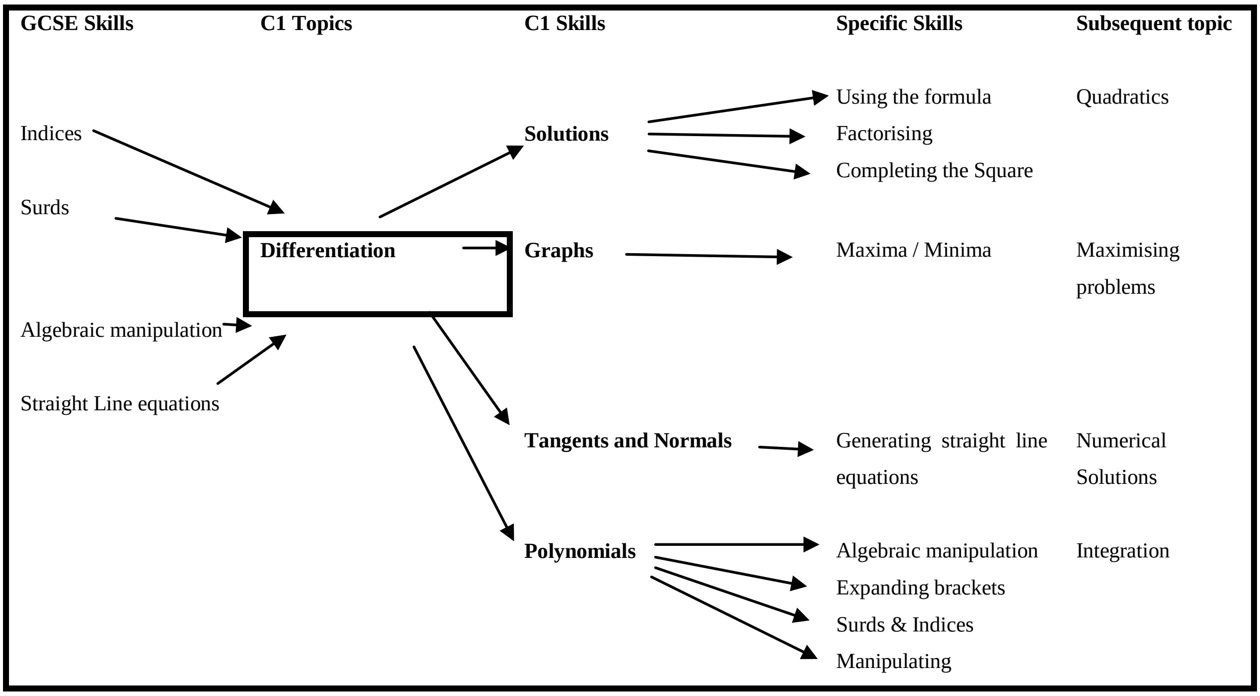


Table 4.8: A content course dictionary for differentiation in Core Maths 1 (C1).

4.5 Conclusion and future direction

Only when mal-rules are fully understood (and re-evaluated regularly) can they be encoded successfully into questions to provide the required discrimination within those questions, and equally provide useful feedback to students once they select the distracter thus encoded. The successful categorisation of such mistakes leading to refinement of questions, learning outcomes and student profiling for teaching and assessment will be of equal importance too, as developed through the SOLO error taxonomy above.

The observation that most errors (as depicted in table 4.6) for the differentiation of a polynomial revolve around uni-structural solutions, could lead us to hypothesise most mal-rules for similar multi-structural problems (for such questions or topics where we had little or no evidence or student errors). In these cases, we would only need to break down a tested skill into its component processes, a failure of each of which would lead to a new mal-rule. The scope for collection of student errors across all topics is beyond most researchers, but the more accurate the mal-rules, the more accurate the diagnosis and feedback. With accurate error accumulation across all topic areas, we can build up specific tailor-made mal-rule banks for import into our question banks.

Chapter 5 – A new diagnostic test

5.1: Setting up a diagnostic test

This chapter presents a simple resource that allows the teacher to accurately diagnose some key weak areas that I see as integral before embarking on an AS course. This will allow students the chance to go off and practise those skills themselves with as much help as they / the teacher wants before the course starts. It should provide ideas (and easy adaptation) for teachers to come up with more advanced or parallel diagnostic tests, which can also be used by students for revision. Moreover, for higher aiming higher-tier students, this AS diagnostic test would double up as a formative GCSE algebra revision test.

The previous chapters have outlined the need for accurate diagnosis, and even more to the discovery of mal-rules and associated student errors through evidence from student work alongside hypothesised errors based on working out which key processes are required to answer each question.

The diagnostic test presented in this chapter is a versatile multiple-choice question Excel worksheet, full of distracters based on mal-rules (with answers and feedback available to the students). The rationale behind multiple-choice questions here is four-fold:

- It is a new style to the students, so they see immediately that there is a different style of questions set for sixth-form mathematics
- The multiple-choice answers are coded with mal-rules – common and likely mistakes as alternative answers to the correct ones, to clearly identify where students go wrong and also to encourage them not to guess (a random answer being: “none of the above”).
- It is very easy for the students / teachers to mark and also to redo as many times as needed.

- It has been set up easily as a non-calculator paper, testing obvious mental maths skills, in preparation for the non-calculator C1 paper.

The previous chapter showed just how many possible errors from uni-structural questions were possible, for a skill of differentiation of polynomials. Given that calculus underpins much future mathematical work, and algebraic skills underpin some of those skills required to be successful at calculus questions, it would seem prudent to start our design of tests using algebraic questions, covering a wide range of skills, so that each question in a test could look not only at different skills, but pose the question to need a multi-structural response. Following the basic GCSE syllabus overlap with the AS course (see figure 5.1), the University of Glasgow diagnostic tests, **Gibson (2004)**, and University of Denver entry profile survey: (http://www.du.edu/nsm/departments/mathematics/coursesandadvising/mathproficiencysurvey.html), the following skills sets were written into this diagnostic test (and the test is shown in full in Appendix 1):

- Simplifying expressions with indices
- Solving linear equations (with brackets)
- Factorising linear expressions
- Simplifying algebraic expressions (including algebraic fraction)
- Multiplying and simplifying / factorising algebraic fractions
- Multiplying our quadratic brackets
- Solving linear inequalities
- Evaluating expressions containing numeric fractional terms
- Solving linear equations with fractional coefficients
- Solving quadratic equations (including word-based algebraic problems)
- Algebra – solving equations and simplifying using the modulus function

I have included a small final section on the modulus function – two simple questions based on evaluation and solving of a simple linear modulus equation. The rationale behind such an inclusion is two-fold:

- It should be a new concept to many of them, albeit readily understood within a few minutes teaching time before the assessment.
- It serves as a control section, i.e. should serve as a great discriminator on these assessments, and allows a measure of control and necessary comparison of facility value and discrimination via other questions.

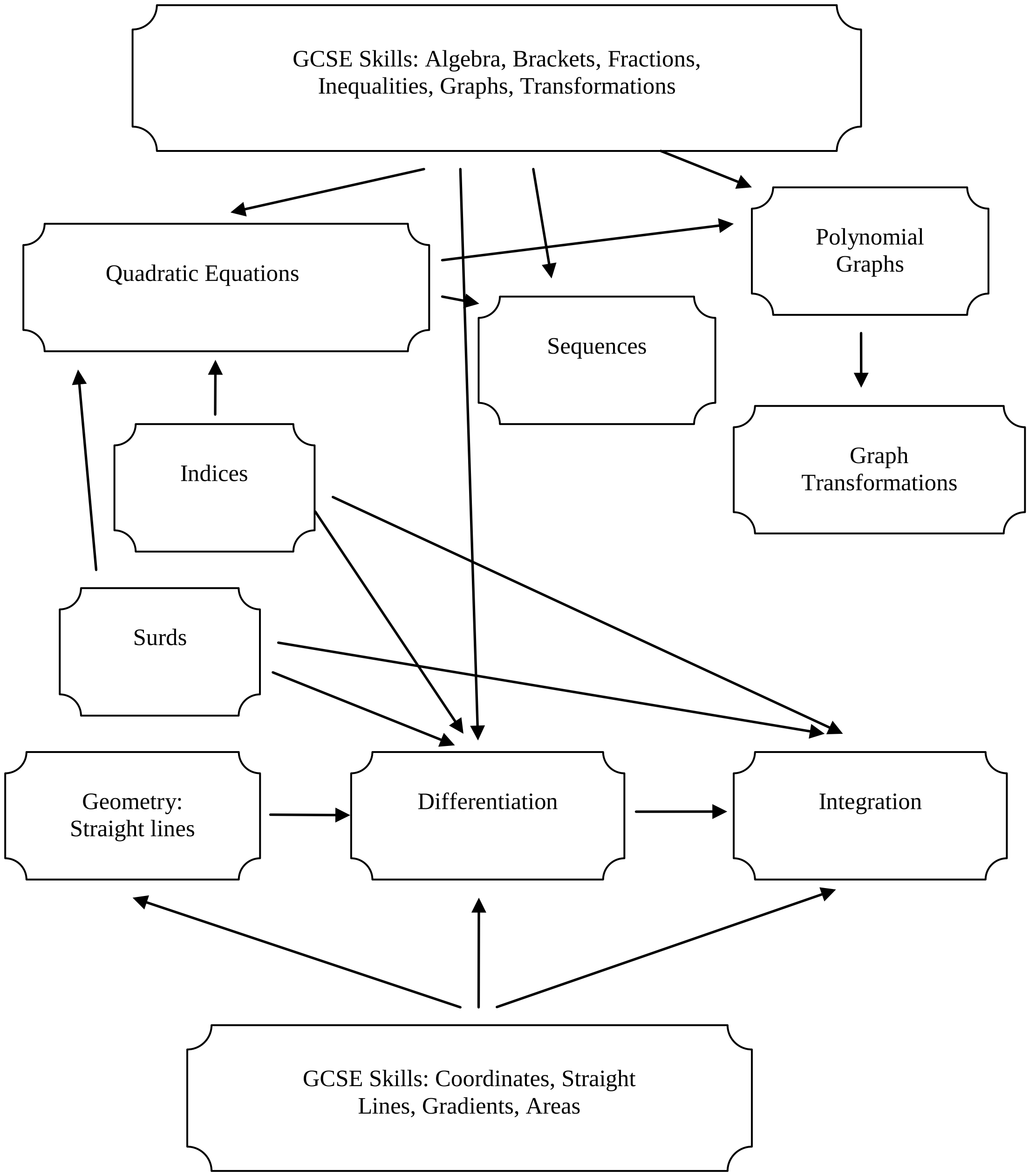


Figure 5.1: Topic dependency spider diagram for C1

5.2 Question design and mal-rules

In building up any multiple-choice test, we must recognise that the questions are testing skills, not topics, and the same goes with designing mal-rules, if we are to devise effective ‘distracters’ (wrong answers) to questions, **Greenhow (2000).** While there has been much recent work completed on understanding students’ misconceptions in foundation algebra courses, my work in chapter 3 has equally shown that in designing useful distracters for this present test needs much evidential help, **Gill & Greenhow (2006).**

I chose to use a sample of 25 students from two independent centres who sat the initial paper-based test of 16 questions between June and September of 2003. The multi-structural questions demanded a final numerical response or final algebraic simplification and aside from the initial questions, I offered them no other explicit cues to determine their strategy or structure of response. From these solution sets, I was able to categorise many of the common misconceptions, pre-structural mistakes and uni-structural mistakes on each of the test questions and so start to hypothesise the best mal-rules and hence the most useful diagnostic distracters to build in to the multiple-choice answer scheme, **Greenhow (1996).**

Some of the pre-structural responses included clear mis-reading of questions, mis-understanding terms (factorise instead of solve), sign errors, and wrongly assuming the question was like another sort of problem they knew how to solve. The uni-structural errors involved basic calculation slips, and even adding in invisible brackets or ignoring visible ones. I include two analyses of the questions (Q5 and Q15) on the test to show how I arrived at the mal-rules:

Question 5: Simplify: $5x-[2x+4(x-3)]$

Despite appearances, this was the hardest question on the paper (lowest facility value of 0.24), with only 6 correct responses of the cohort of 25 students. The most common mistake was to insert another bracket and treat it as if it would result in a quadratic expression:

$$\left[2x-3(4x-1)\right]=\left[(2x-3)(4x-1)\right]$$

Solution	-x+12	-x-12	7x-12	3x-8	5x-6x+12	-x+3	$2x^2$+7x-12	$30(x^2$-2x)	x^5-60	Blank
Frequency	6	4	2	1	1	1	4	3	1	2
Error	Nil	US	US	PS	US	US	PS	PS	PS	PS
Mistake	Correct	Misread + for -	Misread + for -	Wrong factorised x term	Not simplified	Ignored 4 multiply by -3	Multiplied out inner brackets (as above)	5x has multiplied not - the inside term	Unclear pre-structural mistake	No idea

Figure 5.2: Initial results for pilot diagnostic test Q5 without mal-rules (25 students): PS = Pre-structural mistake, US = Uni-structural mistake

One answer with a quadratic is understandable, but mistakes resulting in higher powers do show a serious pre-structural issue for students sitting problem like this. However, it is important to set more problems of this nature because, like question 2, brackets inside brackets do seem to cause many problems for students as they require many processes to be correct en route to the correct final response and yield many types of solutions. To use most of these solutions in a multiple choice answer would be prudent, as many mistakes were being repeated. I used the 4 most common mistakes, making sure two of them were pre-structural and two were uni-structural, alongside the box “None of these”, another pre-structural response. When building a repeatable test like this, it is important that the errors are not only categorised, but clearly identified, so that a generic mal-rule is encodable into a question when we change the coefficients in the question.

Thus my suggested answers are (including re-writing –x+12 as 12-x):

A)	B)	C)	D)	E)	F)
7x-12	$2x^2$+7x-12	-x+3	$30(x^2$-2x)	12-x	None of these

Figure 5.3: Suggested multiple-choice answers for Q5 on the pilot AS diagnostic test

Question 15: Solve for x: $x^2 - 6x + 5 = 0$

While different to Question 5, this was designed as a simple quadratic question that should have been accessible for any GCSE student, as I aimed to test rudimentary understanding of the twin skills of solving quadratics and understanding the nature of basic polynomials. Such a question would allow the student the choice of three or four processes or structures to find *x*. Pleasingly, the vast number of solutions were correct (facility of 0.64), but there were still a handful who made uni-structural mistakes when factorising the expression / using the quadratic formula or indeed pre-structural mistakes when not solving it at all. Designing mal-rules based on evidence becomes even more reliant on the nature of the pre-structural and uni-structural mistakes when the facility value is high.

Solution	1, 5	1, 2	-1, -5	(*x*-1)(*x*-5)	(*x*-2)(*x*-3)	Blank
Frequency	16	1	1	1	1	5
Error	PS	US	US	US	PS	PS
Mistake	Correct	Mistake on formula / factorising	Mistake on formula / factorising	Factorised only (albeit correctly)	Mis-factorised and not solved	

Figure 5.4: Initial results for AS diagnostic test Q15 without mal-rules (25 students): PS = Pre-structural mistake, US = Uni-structural mistake

The main tested skill in this question is intended to be solving the quadratic equation, not specifically on the definition of the word ‘solve’, so I propose only using factor pairs in the main solutions, rather than leaving answers in a factorised form. The logic behind this goes as follows: as solutions indicating factors will be a) a giveaway of part of the solution and b) likely to draw students towards the other solutions when they see the contrast, the solution set should, in my view, just include root pairs. Although this gets round one hurdle, it does allow a substitution checking method to be used, yet I feel this is a price worth paying to get the students to think about the processing involved in solving equations of this form. I thus chose pairs which are easily generated by mistakes with signs in factorising and in the formula itself. The nature of these choices is that two of them would clearly be the result of one sign error in the factorising process (uni-structural mistake), while two other distracters would need more than one error to be made with signs in factorising (or using the quadratic formula) which I would deem a pre-structural mistake (alongside the "none of these" choice). Figure 5.5 below: Suggested multiple-choice answers - Q15 on the pilot AS diagnostic test:

A)	B)	C)	D)	E)	F)
1, 2	1, 5	-1, -5	1, -5	1, -2	None of these

5.3 Overall Results

AS diagnostic test scores (101 students)

Summary	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Correct	69	51	92	64	57	45	47	48	69	51
US	12	27	8	12	18	21	10	11	9	22
Misread (US)	0	0	0	0	3	0	1	0	0	0
PS	20	21	1	22	23	32	29	29	20	19
Blank (PS)	0	2	0	3	0	3	14	13	3	9
Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Facility	0.683	0.505	0.911	0.634	0.564	0.45	0.465	0.475	0.683	0.505
Difficulty rank	10	5	16	9	7	2	3	4	11	6

Summary	**11**	**12**	**13**	**14**	**15**	**16**	**All Q1-16**	**%**
Correct	89	75	22	54	67	67	**967**	**59.84**
US	6	9	36	12	12	11	**236**	**14.6**
Misread (US)	0	5	0	0	0	0	**9**	**0.557**
PS	3	8	37	18	11	9	**302**	**18.69**
Blank (PS)	3	4	6	17	11	14	**102**	**6.312**
							1616	**100**
Question	**11**	**12**	**13**	**14**	**15**	**16**		
Facility	0.881	0.74	0.218	0.53	0.663	0.663		
Difficulty rank	15	14	1	8	12	13		

Table 5.6 (overleaf): Overall scores on AS diagnostic test by question, subdivided by PS = pre-structural mistake, US = Uni-structural mistake.

The table (5.6) overleaf shows the results gained from the first 101 students who sat the AS multiple-choice diagnostic test between 2004 and 2007 with the test complete with distracters. As can be clearly seen, there is one standout question with very low facility (Q13), and two questions (Q3 and Q11) for which the facility was around 0.9, i.e. almost all students achieved the correct answer. None of these questions would serve as useful discriminators. To achieve a roughly average question facility of around 0.60 is a level consistent with any diagnostic tests (and hence much more useful for discrimination).

Within the test summary, I have subdivided the uni-structural mistakes into those where I can see a clear mis-read of the question, and the majority which have been part correct processes. Within the pre-structural mistakes, I have subdivided the categories into blank (where I have no clue which if not all of the processes where outside of the students' domain of knowledge) or other pre-structural mistakes, where the question yielded a multi-structural mistake in their response.

In short, a pre-structural or uni-structural mistake is really too overarching to be a useful classification for designing mal-rules on its own, so the fact that there are roughly the same numbers (and proportions) of uni-structural and pre-structural mistakes gives weight to writing an equal number of distracters that cover both levels of error. This itself, coupled with the evidence gained from the first 25 students to sit a paper-based (non-objective) version of the test, should help us feel confident in our selection of evidence-based mal-rules.

As detailed earlier, our control section (modulus questions) yielded a facility of 0.564 (from table 5.7 below), which sat in the middle of the topic areas on difficulty. This begs the question: why are some questions harder than others or indeed harder than the control set? In some cases, it is the simple nature of number of processes or operations needed to complete the question (so each time adding to the opportunities to make a uni-structural mistake). While in other cases, the concept depth is much more than other questions (i.e. Q8 when simplifying an algebraic fraction), and there are fewer if any explicit cues to start off the question. The other reason why some questions have lower facility values is down to the question reusability – and pedagogy. Have I written the wrong mal-rules or written them in such a way as to overly distract a student from the correct answer by phrasing it in an unfamiliar way? I have only used one question on each of inequalities and numerical fractions, so it is not easy to draw conclusions from a one-question section. Even if I were to use a much greater length test to make generalisations, I suspect that similar difficulties would show up from my own experience.

Summary Q areas	Modulus (Q1,6)	Indices (Q9,10)	Equation (Q12,14)	Brackets (Q5,11)	Algebraic Fraction (Q4,7,8)	Inequality (Q13)	Numerical Fraction (Q2)
Correct	114	120	263	146	159	22	51
US	33	31	44	24	33	36	27
Misread (US)	0	0	5	3	1	0	0
PS	52	39	46	26	80	37	21
Blank (PS)	3	12	46	3	30	6	2
Facility value	0.564	0.594	0.651	0.723	0.525	0.22	0.505
Difficulty rank	4	5	6	7	3	1	2

Table 5.7: Summary of AS diagnostic test results by question area.

This test has the benefit of fitting neatly into a 40 minute window (regularly the length of single lessons in secondary schools), as the timescale was developed in conjunction with the length of time needed to complete the paper-based test by the first 25 students. Getting the length of test wrong (for the time available), is never ideal, and difficult to plan without pre-testing of students as a pilot scheme, **Gill & Greenhow (2006).**

Where the test falls down here is that it is non-reusable, paper-based and students must all sit exactly the same test (thus allowing for potential direct copying or just collusion). It also doesn't yield individual feedback easily, as the tests need to be marked, and student selections need to be matched to a pre-prepared paper-based answer sheet showing which types of mistakes they have made based on their wrong choices. Such diagnostic tests become really useful only when they become reusable objects, **Greenhow (1996)**, and so I turned this paper-based test on these questions into an Excel worksheet, with random parameters in the questions and associated random parameters in the answers based on the same mal-rules.

It works especially well when the mal-rules in the answers can be coded to produce exact answers the students might (in error) achieve themselves. Students tend to be drawn to integer solutions or "nice numbers" e.g. ½ rather than more random decimal or surd answers, and such distracters can be easily discarded by students keen to take a short-cut. Thus, some questions are reverse engineered, i.e. I started by generating the integer solutions using mal-rules of a generic question type (e.g. Q15 when solving a quadratic equation), and reversed the processes to lead to that specific question. Using Excel's functionality, if we start with the generation of random parameters as solutions, we can codify the inverse function that yields the question. Each question will produce a minimum of 5 and maximum of 1092 realisations. Thus millions of pedagogically and algebraically equivalent tests can be generated **Gill & Greenhow (2006)**. There is also a column, F, "none of these", and only Q14 is set fulfil that criteria for this first iteration.

AS diagnostic test questions (multi use version)

Q1	Evaluate	$\lvert -19 + 3 \rvert$
Q2	Evaluate:	$\frac{-4}{5}\left(\frac{2}{3} - \frac{1}{5}\left(\frac{3}{4} - \frac{1}{2}\right)\right)$
Q3	Factorise	$42x - 7$
Q4	Simplify	$\frac{4}{x} - \frac{12}{y}$
Q5	Simplify	$2x - (5x + 11(x - 4))$
Q6	Solve	$\lvert x + 5 \rvert = 8$
Q7	Simplify	$\frac{x^2 - 4}{5x} \cdot \frac{25}{5x - 10}$
Q8	Simplify	$\frac{6x - 54}{x^2 - 6x - 27}$
Q9	Simplify	$(-7a^3b^5)(4a^6b^3)$
Q10	Simplify	$\left(\frac{4x^8y^4}{8x^3y^7}\right)^3$

Figure 5.8: Excel version (dynamic) of the AS diagnostic test (Q1 – 10 only).

5.4 Solutions and Feedback

What is also needed are full answers for the paper-based test (so students can take that away as practice worked solutions) and also a feedback / answer sheet for teachers to use when looking through students' work with them. Such a test can be sat with questions online and students submitting their responses via paper-based answer sheet. The nature of the feedback, and its efficacy is thus very important when using this as a reusable diagnostic tool, especially is the feedback can be delivered onscreen. A sample of the Excel feedback screenshots is shown in Appendix 2, and it is already clear that the formatting issues with Excel leave feedback challenges.

Answers

Q	A	B	C	D	E
1	-16	57	22	16	-22
2	-0.58	-0.49	-0.33	0.5	-0.57
3	x= 0.167	$7(6x-1)$	$7(6x-7)$	$6(7x-1)$	$7x-1$
4	$\frac{4y-12x}{xy}$	$4y-12x$	$\frac{-8}{(x-y)}$	$12x-4y$	$\frac{12y-4x}{xy}$
5	$8x-44$	$5x^2+14x-44$	$-4x-20$	$32x^2-88x$	$44-14x$
6	$x=-3$	$x=3$ or 13	$x=3$	$x=3$ or -3	$x=3$ or -13
7	$x^2-25x+10$	$\frac{-3}{x}$	$\frac{x}{x+2}$	$x^2-25x+10$	$\frac{x+2}{x}$
8	$\frac{6}{x+4}$	$\frac{6x}{x+3}$	$\frac{6}{x+3}$	$-6x-54$	$\frac{-27}{x^2}$
9	$-28a^{18}b^{8}$	$-28a^{18}b^{15}$	$-28a^{18}b^{18}$	$-28a^{9}b^{15}$	$-28a^{9}b^{8}$
10	$\left(\frac{x^5}{2y^3}\right)^3$	$\left(\frac{x^{33}}{8y^9}\right)$	$\left(\frac{x^5}{8y^9}\right)$	$\left(\frac{x^{15}}{8y^3}\right)$	$\left(\frac{x^{15}}{8y^9}\right)$

Figure 5.9 – Answer sheet for columns A to E (Q1-10) of the AS diagnostic test.

There is need for careful consideration here about when students are mature enough to be able to plug gaps themselves, as indicated earlier on, and also the whole realm of feedback use for students who are keen to build up skill levels. **Gill & Greenhow (2006)** explored how students were often being distracted by their excellent and detailed feedback and not moving quickly onto sufficient numbers of questions to make useful practice gains. For those unable to work through this e-learning for themselves, at least knowing weak areas or weak skills will allow those who want to bridge the gap to do so with their teachers and learning mentors.

5.5 Further discussion

In conclusion of this chapter, I present some further analysis of results gained by students taking the AS diagnostic test and their subsequent performances on AS exams (especially C1 papers). It is no surprise that there is a reasonably good correlation between scores on the diagnostic test and subsequent scores throughout their sixth-form courses. This positive result does show the power of diagnosis.

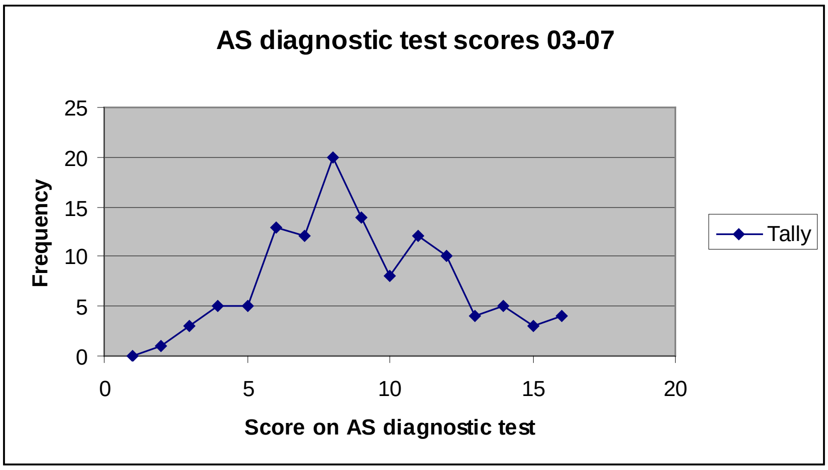


Figure 5.10 – Score distribution for the 126 students who have sat the AS diagnostic test, 2003-2007

As can be seen clearly, there is a vague bell-shaped distribution to the results gained by students sitting the diagnostic test, and it is noted that students scoring 5 or less out of 16 on the test are highly unlikely to achieve a high grade at AS maths. There were a couple of exceptions (those who scored 5/16 on the test and then an A grade at AS maths), but all of those who scored a grade C or less on the AS maths exams had scored 8 or less on the diagnostic test.

Rank Correlation	Diagnostic score v C1 score	0.551869
	Diagnostic score v AS total score	0.649064
	Diagnostic score v A-Level total score	0.514488
	C1 score v AS total score	0.877785

Table 5.11 – Rank correlation results for Diagnostic test score v other scores

Clearly students doing well on C1 went on to do well at AS overall (correlation of 0.88), and those who did well at the diagnostic test had a good ranking agreement with the overall AS score (correlation of 0.65), but the comparisons between diagnostic test score and C1 or A-level score were less convincing (0.51-0.55).

One reason why students went on to do well at AS maths overall is through using the feedback well from the diagnostic test. Those who built on their domains of knowledge as well as ability to understand and apply the knowledge best of all will have achieved the highest grades overall. It is the nature of support given to students to improve skills and cut out uni-structural mistakes that yields greater potential final scores, and schools and colleges should be motivated to offer an ever-increased and tailor-made set of courses and help for students. The efficacy of such support, whether through e-learning, maths clinics, additional teaching time, one-to-one tutor support or the like is something that needs investigating further.

5.6 Retakes and retests

When the first two cohorts of students were retested (one week later), using the multiple-choice answer sheets (and additional time spent practising processes and skills in the classroom), the average marks improved from 7.0 to 10.0 out of 16. The correlation between the average of retest score and overall AS score improved to 0.76. This gives a good indication of the use of the follow-up retest to help us make predictions of final AS grades, and to underpin future skill development with further time spent consolidating basic skills needed in transition. Once on the C1 course, and successfully completing a remedial course in specific key GCSE skills that were lacking, students should have the best chance of realising their potential upper grades. Clearly the better the diagnosis and better remedial action available to students, the better the final result.

This is exemplified through a further investigation of the nature of retests through retakes on C1 papers (and differentiation skills) based on data from Edexcel’s ResultsPlus format and using data from chapter 3. Of the 65 retake papers sat by students, all but 9 of them either improved their percentage score on differentiation or achieved the same (100%) understanding of the topic 2nd time round. The average improvement was 12% on this topic alone from the 1st to the 2nd retake. However, for those who sat the paper for a 3rd time, 4 improved, 8 stayed the same and 5 went down (an average of 1 %). Clearly this is a topic where the weaker candidates are learning to some extent from their mistakes, but the very weakest candidates are not learning enough from their returned exam scripts, diagnosis of errors or revision work to make the inroads on this particular topic.

A computer-based learning and assessment scheme is therefore likely to yield the sort of improvement that they need by the 3rd attempt at such a paper, and shows equally that learning for retests needs to be very specifically focussed and supported, as students are likely to assume that they understood a skill previously, so they can handle any questions on that topic or skill next time round. Revision of students' own perceived strengths is often discarded at the expense of time spent on students' own perceived weaknesses.

5.7 Issues with encoding mal-rules

One target question at the start of the thesis was whether students made the same mistakes on a paper-based test as onscreen.

On some of the questions from the diagnostic test, there is danger that students will guess the correct response by the similarity of neighbouring possible choices, or indeed rule out distracters by the difference from other possible solutions. If a question changes facility value greatly between a paper-based test and onscreen test, I have to ask myself: have I written the wrong mal-rules or written them in such a way as to overly distract a student from the correct answer by phrasing it in an unfamiliar way?

In Q5: the danger of adding both $-x+12$ and $-x-12$ to our multiple-choice solution set is one whereby students could guess the correct answer by similarity, so as such, I have ignored one mal-rule that would lead to a solution too close in stature to the final correct response. The pre-structural mistake involving (much) higher powers, is another one I have also ignored, as such a response could intuitively look so different to the other responses that it could overly distract a student from considering it. In Q15, however, I chose pairs of integers which are easily generated by mistakes with signs in factorising and in the formula itself, rather than encode some of the more popular mal-rules as distracters. I mentioned earlier that students were often drawn in some way to intuitive solutions, be they integer solutions or numbers like "½" rather than more random answers, and so distracters must be worked out carefully to stop short-cuts.

Having written the multiple choice answers to these questions, and the next 100 students sat the test, Q5 improved from being the hardest question to become only the 7th most difficult question, whereas Q15 had no change on the difficulty rank (12th out of 16) between paper-based numerical response format and multiple-choice response format. Clearly the nature of selection of mal-rules has changed the facility of Q5 considerably. There is a question of analysis in the future on the role mal-rules play in guiding students away from or towards specific answers, **Greenhow (2009)**. When looking further into this change on Q5, the pre-structural mistake of including an invisible bracket caught out many students with a paper-based response format but, when offered as a distracter, it was clearly seen as wrong, and hardly selected at all. Our preliminary investigation here suggests that the nature of some multiple-choice answers will clearly cause the students not to follow that same process.

Using the templates from **Greenhow at al. (2003)** for diagnostic question design of multiple-choice questions, we can include a null ("none of the above" or "I don't know") choice as well. They showed the need for a null choice answer box in Mathletics, which was the correct answer an author prescribed probability (usually 0.1), to make sure students fully considered every option and so didn't guess based on partial knowledge. It also aimed to dissuade students from attempting to eliminate all of the answers bar one in case some distracters proved to be too far-fetched (i.e.

very unlike any of the others) and ruled themselves out by observation (as seen with the transition from paper-based to onscreen test Q5 above). In the first two cohorts of students (30 students) who sat the multiple-choice version of the test, the number of correct choices of the “none of these” answers was around 66%, which shows that it was neither overly nor under-used.

5.8 Limitations of model

Our diagnostic test can be sat with questions online and students submitting their responses via paper-based answer sheet. A discussion point for the future would involve selection of responses using a macro-program embedded within the Excel worksheet that would generate the summary / feedback sheet on-screen after the end of the test, and record the students’ marks for summative assessment purposes. In the short term, the many varied CAA and e-learning resources, like Mathletics, will serve us well as an evolution of diagnostic tool available to students and teachers.

Chapter 6 - An analysis of common errors and their associated diagnosis from "results plus" feedback of Core maths 1 scripts.

6.1: A study of Core Maths 1 exam script results via ResultsPlus

The motivations for studying a vast number (374) of Core Maths 1 (C1) scripts are:

- To diagnose generic weak area (skills and topics) from students within the same centre, so to impact teaching for future students and retake preparation for present students.
- To learn how best to structure revision aids and practice exam papers, in preparation of sitting such C1 exams in future.
- To diagnose specific weak areas (skills and topics) for students who wish to retake the module, as well as for those wishing to sit further modules in core maths.
- To assist in writing further CAA revision and examination materials using common skill requirements, and identification of key skills that need to be understood from A-Level before students at tertiary level can build on this foundation level maths.
- To further categorise common errors and misconceptions and derive common mal-rules in order to help develop future e-assessment and e-learning diagnostic and formative assessment tools.

Edexcel (via : www.edexcel.com/resultsplus) offers immediate post results feedback to centres (and candidates) via results files in Excel, which break down the 75 mark core maths 1 paper on a mark by mark basis, providing binary feedback (as in table 1 below) for each candidate on the same mark by mark basis. The screenshot below is the first 21 marks of the paper from January 2008, for the 12 candidates who sat it via our centre. Candidate names have been replaced with letters A – L.

ResultsPlu
s

EXAMINATION GCE JANUARY 2008
SUBJECT 6663 CORE MATHEMATICS
1

CANDIDA	A	A			a	b	b	A	A	M	M	aA	aA	aM	aM	bA	bA	b	aB	aB	bA
TE	1	2	B	M	B	A	M	1	2	1	2	1	2	1	2	1	2	M	1	2	1
A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
B	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1
C	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1
G	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
H	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	1	1	1	0	1	1
I	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
J	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
K	1	1	1	1	1	1	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1
L	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1

Table 6.1: Screenshot of the results file (CSV) from Edexcel's results plus facility, January 2008 C1 module scores.

The headers for each column refer to the following examiners' shorthands:

M1 or aM1	Method mark allocated for stage 1 of the process in a one part question (aM1 is the mark allocated to the 1st stage in part (a) of the multi-part question.
M2 or aM2	Method marks allocated for stage 2 of the process in a one part question (aM2 are the marks allocated to the 2nd stage in part (a) of the multi-part question.
A1 or aA1 or bA1 or bA	Answer mark allocated for the question following on from Method (M1) mark previously achieved (aA1 or bA1 / bA are the answer marks allocated to in part (a) or (b) of the multi-part question).
A2 or aA2 or bA2	Answer marks allocated for the question following on from Method (M1) mark previously achieved (aA2 or bA2 are the answer marks allocated to in part (a) or (b) of the multi-part question).

B or aB1 or aB2	Answer mark allocated for the question without any derived method shown (aB1 or aB2 are the answer only marks allocated to in part (a) of the multi-part question.)
M or bM	Method mark allocated for stage 1 of the process in a one part question (bM is the mark allocated to the 1st stage in part (b) of the multi-part question.

Table 6.2: Categorisation of examiners' feedback marks for Edexcel maths scripts.

The disadvantage of such a results file is that it clearly doesn't give us the exact candidate answers for us to analyse exactly where students have gone wrong. In an ideal world, we would wish to access all 374 student scripts, and to look through similar questions or skill tags to seek out common student errors and misconceptions. To attempt this by hand is impossibly time consuming and exceptionally expensive (by a factor of over 3 times the standard entry costs per module).

Results Plus answer files' real advantage over the raw scripts is that we can easily analyse trends over one exam session (given sufficient candidates), as the data is exportable to Excel format. By categorising each mark / question with suitable tags, we can also seek collective weak areas / skills / topics, and we can infer general trends temporally across exam sessions to assess whether weak areas / skills / topics continue to trouble students that have undergone almost identical preparation year by year.

We can also identify from these files those candidates who have sat the same paper more than once (and of the 374 answer files, I have identified 17 candidates who have sat this same module 3 times, and a further 130 students who have sat the paper twice), results summarised at the end of the previous chapter. From these repeat performances, we can work out the efficacy of the students' revision for the retake(s) have been by topic area or skill, and also pick up trends in continued weak area / topics / skills.

I have already discussed the importance of understanding student errors and how they can lead to mal-rules, and equally categorising them using the SOLO taxonomy.

In this analysis, I will show how mal-rules can be inferred, as well as topic / skill deficiencies either by individual feedback or by cohort from raw marks. Subsequent chapters will allow an alternative focus on errors, mal-rules and diagnosis from CAA data and e-assessment.

6.2: Results by topic for core maths 1

Over a period of 3 years (and 6 exam sessions from June 2007 through to and including January 2010), covering 374 candidate scripts, I have analysed the results files from Edexcel. The first stage of analysis of these results was to take the paper apart and carefully allocate each mark to a topic, sub-topic and skill (s) set.

ResultsPlus	Ind	Int	Frac	Transf	Differ	Quadr	Polyn	AP	Lines	Inequ	Formula
EXAMINATION GCE JANUARY 2008	19	10	4	7	14	20	13	7	11	7	22
SUBJECT 6663 CORE MATHEMATICS 1											
Question	1	1	1	1	2	2	2	3	3	3	3
Mark type	A	A	B	M	B	A	M	A	A	M	M
Topic	Int	Int	Int	Int	Ind	Ind	Ind	Surd	Surd	Surd	Surd
Sub-Topic	Ind	Ind	Ind	Ind	Frac	Frac	Frac	Quadr	Quadr	Quadr	Quadr
Skill	Simp	Simp	Iden	Manip	Iden	Simp	Iden	Manip	Simp	Iden	Manip
Skill 2			Simp	Simp	Eval	Eval	Simp	Eval	Eval	Manip	Simp
Other skill / topic	Indef	Indef	Indef	Indef	Frac 1/a	Frac 1/a	Frac 1/a				
CANDIDATE	1A1	1A2	1B	1M	2aB	2bA	2bM	3A1	3A2	3M1	3M2
A	1	1	1	1	1	1	1	1	1	1	1
B	1	1	1	1	1	1	0	1	1	1	1
C	1	1	1	1	1	1	0	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1
E	1	1	1	1	1	1	0	1	1	1	1
F	1	1	1	1	1	1	0	1	1	0	0
G	1	1	1	0	1	1	1	1	1	1	1
H	1	1	1	1	1	0	0	1	0	0	0
I	1	1	1	1	1	0	0	1	1	1	1
J	1	1	1	1	1	1	0	1	1	1	1
K	1	1	1	1	1	1	0	1	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1

Table 6.3: Results plus file from January 2008 (C1) annotated with skill and topic headers.

The first four columns refer to the question 1: Find $\int (3x^2 + 4x^5 - 7)\ \mathrm{d}x$. [4]

Question	1	1	1	1	Key to short form of terms
Mark type	A	A	B	M	Int = Integration
Topic	Integrate	Integrate	Integrate	Integrate	Ind = Indices
Sub-Topic	Indices	Indices	Indices	Indices	Simp = Simplify
Skill	Simplify	Simplify	Identify	Manipulate	Iden = Identify
Skill 2			Simplify	Simplify	Manip = Manipulate
Other skill / topic	Indefinite	Indefinite	Indefinite	Indefinite	Indef = Indefinite

Table 6.4: Sub-categorisation of marks on Q1 of January 2008 C1 paper by topic and skill

The mark scheme for this paper divides up the marks for each question and part into 3 distinct areas: M (method), A (answer derived from a previous method), and B (answer derived without prior method, in this case the "+c"), as detailed in table 6.2. I have used a number of short forms (as shown in column 6 in table 6.4) throughout the analysis to simplify character and box sizes on the Excel analysis chart. In question 1 above, we have a question on indefinite Integration (Int), with associated sub-topic: Indices (Ind). The marks are allocated firstly for method (the demonstration of an attempt to correctly integrate one term, for which I have designated the skills: Manipulate and Simplify), and then three marks for correct answers (the first of which requires one correct power term, the second and third for the correct other terms). These subsequent answer marks, I have designated the skills required as: Identification / Remember (Iden) of both the (+c) and also the nature of solutions, and also Simplification (Simp) for the fractions. A fuller description of how skills are allocated per question is carried out later in the chapter.

The "ResultsPlus" spreadsheet tells me clearly which candidates have scored which marks on question 1, and where they lost the marks if at all. In the above table, we can clearly see that only candidate G lost a mark at all on this question, and that for the "B" mark, i.e. his mistake was forgetting the "+C" on his solution (i.e. detailed in the examiners' solutions). All other candidates had scored the correct answer in its entirety. Interestingly, should we use a similar question in a multiple choice objective test, with associated mal-rule (forgetting the "+ C"), then I could hypothesise that candidates are likely to rule out this answer when they see a clutch of other solutions that do include the "+C" term, as they are likely to jog the students' memories of the correct form of the integrand, so such a skill would need to be tested using other objective formats. From a teaching perspective, it shows one major deficiency on that question can become a useful revision tip.

Question Number	**Scheme**	**Marks**
1.	$x^3 + \frac{2}{3}x^6 - 7x + c$	M1 A1 A1 B1 (4) **(4 marks)**

Figure 6.5: The mark scheme for Q1 of January 2008 Core Maths 1

6.3 Topic Analysis

What I set about constructing here was a fuller description of each part of every question and the topics and skills required to score every mark on the paper. I subdivided each mark into five categories: Topic, sub-topic, Skill 1 required, Skill 2 required and other information to help categorise the mark concerned. Within this analysis, I was able to carefully allocate every mark to a topic or two and also to one or more skills that were required to be demonstrated. Too few topics / skills and the analysis would be meaningless to the candidate, and too many topics / skills and there would be difficulty classifying the marks as well as usefully highlighting patterns in errors or learning gaps for the candidates.

Topics considered were taken from the June 2007 through to January 2010 papers (source: Edexcel Core Maths 1 Specification, www.edexcel.com):

Indices	Integration	Area	Transformations	Differentiation	Quadratics	Polynomials
Lines	Inequalities	Formula	Sequences	Pythagoras	Arithmetic Progressions	Surds

Figure 6.6: Topic sections for C1 papers

For each mark, the main topic of that part of the question was selected from the above list, and the sub-topic was also added to the spreadsheet if appropriate. For example, question 1 of the January 2008 paper (as depicted above) was an integration question (main topic), with secondary emphasis on indices (sub-topic), and thus all 4 marks were allocated to those topics. Question 2 (June 2008) was a Polynomial topic question ("Polyn"), which also required knowledge of Quadratics ("Quadr"), and related to the study of Cubic equations. So I built up spreadsheets for each of the 6 exam papers with all 75 marks allocated to one or (mainly) two topics, which yielded roughly 130-140 topic allocations per paper, as some questions fell overlapping two topics, and others only one.

One aim being to total up how many marks there were for each topic, and then see how many marks a candidate scored on any one topic to produce an analysis chart on a paper by paper basis, before combining the analysis for the whole cohort per paper and then temporally across all 6 papers. For each of the other topic areas, I constructed grids which showed where the candidates scored marks for that topic only, i.e. for the topic "Indices", I include a selection of the columns for this table below in figure 6.7. The score at the end in bold is the percentage achieved on this topic by the candidate.

NAME	Indices	Indices	Indices	…………………	Indices	Indices	Indices	**Indices % score**
A	1	1	1	……………	1	0	0	100
B	1	1	1	……………	1	0	0	100
C	1	1	1	……………	1	0	0	89

Figure 6.7: A selection of results for Indices questions in June 2008 for candidates A - C.

Once the calculations by topic have been completed, a final table for candidate results by topic is prepared for June 2008 C1, a selection of which is displayed in figure 6.8:

15	**Indices**	**Integration**	**Area**	**Transformation**
A	100	100	50	100
B	100	100	100	60
C	89	67	100	60
D	100	100	100	100
E	100	100	100	100
F	56	33	100	80
G	89	78	100	100
H	100	100	50	40
I	100	89	75	100
J	100	100	100	100
K	56	33	100	40
L	56	33	0	60

Figure 6.8: A selection of final results by topic for C1 June 2008 by candidate.

In each case, the result if given as a strict percentage, and where the candidate has scored 70 or more %, I have left the square white, where the score is 50-69 inclusive, they are awarded a yellow box (to indicate a warning notice on this topic or uni-structural responses to questions on this topic), and a red box (likely to make pre-structural mistakes on these questions) for scores below 50%. Such a table can easily be used to identify candidate's own weaknesses (e.g. candidate "L"), and also topic areas that were harder than others / could do with a whole centre improvement (e.g. Integration above).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	98 cands	**AP**		**Differ**	**Formula**	**Frac**	**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**	**Pythagoras**	**Quadratic**	**Sequence**	**Simult**	**Surd**	**Transf**
2	Jun-07	95		95	89		97	90	97	89	90		85	90	82		81
3	Rank	9		10	4		12	7	11	5	8		3	6	2		1
4	Marks	6		13	28		17	6	9	14	12		11	7	7		5
6	12 cands	**AP**		**Differ**	**Formula**	**Frac**	**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**	**Pythagoras**	**Quadratic**	**Sequence**		**Surd**	**Transf**
7	Jan-08	71		90	84	77	88	70	90	91	90	94	83	89		83	85
8	Rank	2		12	6	3	8	1	10	13	11	14	4	9		5	7
9	Marks	7		14	22	4	19	7	10	11	13	3	20	8		5	7
11	107 cands	**AP**	**Area**	**Differ**	**Formula**		**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**	**Pythagoras**	**Quadratic**	**Sequence**			**Transf**
12	Jun-08	97	72	90	97		91	85	85	87	95	100	93	99			88
13	Rank	10	1	6	11		7	3	2	4	9	13	8	12			5
14	Marks	10	4	15	15		6	5	6	24	12	3	15	6			2
16	39 cands	**AP**		**Differ**	**Formula**		**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**		**Quadratic**			**Surd**	**Transf**
17	Jan-09	89		93	92		92	88	96	77	79		78			98	95
18	Rank	5		8	6		7	4	10	1	3		2			11	9
19	Marks	11		13	10		11	7	8	13	7		12			3	6
21	82 cands	**AP**		**Differ**	**Formula**		**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**		**Quadratic**	**Sequence**		**Surd**	**Transf**
22	Jun-09	87		81	86		84	93	90	88	88		85	87		84	
23	Rank	6		1	5		2	11	10	8	8		4	6		2	
24	Marks	8		18	25		17	7	3	13	10		24	7		10	
26	36 cands	**AP**		**Differ**	**Formula**	**Graph**	**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**	**Pythagoras**	**Quadratic**	**Sequence**	**Simult**	**Surd**	**Transf**
27	Jan-10	96		83	92	86	75		69	88	96		83		92	96	88
28	Rank	10		3	8	5	2		1	6	10		3		8	10	6
29	Marks	9		11	17	13	10		7	12	6		34		7	6	7
31	374																
32	candidates																
34		**AP**	**Area**	**Differ**	**Formula**	**Frac**	**Indices**	**Inequality**	**Integration**	**Lines**	**Polyn**	**Pythagoras**	**Quadratic**	**Sequence**	**Simult**	**Surd**	**Transf**
35	Average	89	72	89	90	82	88	85	88	87	90	97	84	91	87	90	87
36	Difficulty rank (av)	11	1	10	13	2	9	4	8	5	12	16	3	15	6	14	7
37	Marks (av)	9	4	14	20	6	13	5	7	15	10	2	19	7	7	6	5
38																	
39	Weighted Average	90	72	88	89	84	88	85	88	87	90	97	84	91	87	89	87

Table 6.9: Summary data for 374 candidates’ C1 performances based on topic analysis (June 2007 to January 2010).

Looking temporally (as in table 6.9), we can see huge variations from paper to paper on topic difficulty, and year to year, and so it is almost impossible to judge the relative merits of each paper unless we look at the average for that topic (percentage wise) and also a weighted mean to take account of the marks available. Over time, we can see very little differences that are meaningful, nor are trends suggesting certain areas of increasing difficulty or decreasing difficulty. However, the summary data in the bottom of this table gives a few very important pointers for teachers preparing students for this exam paper:

- Formula occurs in 10 out of the 75 marks on each paper (each mark has two topics, so we divide the 20 by 2), so a healthy amount of time needs to be dedicated to questions involving the use of formulae, although it is ranked 13th in difficulty, so represents one of the easiest topics.
- Pythagoras Theorem, Sequences and Surds represent the easiest topics for candidates.

- Quadratics (10 marks), Straight lines (8 marks), Differentiation (7 marks), Indices (7 marks), Polynomials (5 marks) and Arithmetic Progressions (5 marks) all feature prominently in the table, and occur in every paper.
- Algebraic fractions, Graphs, Simultaneous equations, while featuring significantly in the syllabus, have failed to turn up often on these recent past papers, and have not attracted that many marks when they have done so.
- Area questions (while only averaging 4 marks on the one time it occurred) represents the most difficult topic by quite some way (72% average scored, compared to 82% for the next most difficult).
- Integration interests me most, as many years it appears to be the easiest part of the paper (rank 10th or 11th), but a couple of years (June 08 and January 10) it is about the hardest question on the paper.

6.4 Skills Analysis

With a very similar approach, I completed the analysis for the C1 results by candidate on a skills basis. A “skill” is an overarching header for a collection of processes that leads to the desired response. For example, to “simplify” an expression requires at least knowledge that similar terms can be collected together, the role of the BIDMAS rule needs to be adhered to, and a consummate understanding of how terms separated by fractions need to be collected differently. While topic classification is unambiguous to all, skill classification could be more subjective. To demonstrate the selection process: I use June 2008 C1 (with associated frequencies for this paper, using mostly two skills per mark):

Manipulate	Identify	Simplify	Factorise	Evaluate	Draw	Substitute	Previous	Show
15	10	29	5	29	6	20	7	5

Table 6.10: Skill set by frequency for June 2008 C1.

In addition, I chose to use a further four skills in the January 2009 C1 paper:

Solve	Derive	Symbol	Fraction
2	2	3	4

These last four skills are of low frequency to the other major 9 skills, and as such are not really indicative of any major failings or success of a candidate. The skill “Factorise” could be incorporated into either “Manipulate” or “Simplify”, and because of this ambiguity, it is left as its own skill set. Its frequency just about justifies this stance too.

Definitions of the skills

Manipulate	An attempt to use a core technique (e.g. differentiation) or algebraic skill (multiplying out brackets) as part of a method to yield an evaluate or expression.
Identify	A recall of formulaic, factual or technical information which generates either a straight answer (B mark), or leads to derivation of an expression or equation for future solving or manipulation.
Simplify	An attempt to take an expression or equation and move / gather terms such that the equation / expression will yield an evaluate or simplificate.
Factorise	An attempt to install brackets separating multiplicative factors involved in an expression or equation.
Evaluate	To take an expression or equation that has been manipulated and simplified such that it yields an evaluate.
Draw	To display graphically the results of an identification or evaluation.
Substitute	An attempt to input data or information from the question, a previous part, or from an identification in order to form a new expression or equation.
Previous	Use of an evaluate from a previous part, an expression or equation from a previous part, or information generated in the calculation in a previous part for future substitution, manipulation, simplification or evaluation.
Show	To successfully demonstrate what was required to be demonstrated (e.g. derivation or proof), having previously manipulated and / or simplified an expression or equation to yield an expression, equation or evaluate as dictated by the question concerned.

Table 6.11: A definition of the skills required for C1 exam papers

As a demonstration of how to allocate the skills for each question, I use the June 2008 C1 paper, and have allocated the following skills to Questions 1 and 2 as below:

1. Find $\int (2+5x^2)\ \mathrm{d}x$. **(3)**

2. Factorise completely $x^3 - 9x$. **(3)**

	1	1	1	2	2	2
	M	A	A	B	M	A
Skill	Manipulate	Simplify	Identify	Factorise	Factorise	Evaluate
Skill 2		Evaluate	Evaluate	Manipulate	Simplify	
Other skill / topic	Indefinite	Indefinite	Indefinite	Cubic	Cubic	Cubic
	Q01M	Q01A1	Q01A2	Q02B	Q02M	Q02A

Figure 6.12: Questions 1, 2 from June 2008 C1 with associated skill set required.

For question 1, the first (method) mark goes for manipulating the integrand (a technique only). The first answer mark requires the candidate to simplify the cubic expression, i.e. to correctly arrive at (hence: simplify and then evaluate skills used): $\frac{5}{3}x^3$

The second answer mark is for identifying the (+c) – a factual recall, and then Evaluating the final two terms: 2x + c

In question 2, the first mark (B) is for factorising out one bracket, hence the factorise and manipulate skills needed to remove one term and leave the rest of the expression correct. The next (method) mark is for a second attempt to factorise into three multiplicative terms, simplifying the answer as one goes. The final answer mark is for evaluating the correct expressions within the brackets from the earlier method.

Subsequent questions have been tackled with the same approach, some yielding two distinct skills, a few marks only one (typically for the final evaluate). A full table of skill allocations is found on the 2nd tab of the "FINAL Jun 08 C1 HS analysis.xls" spreadsheet.

Once the allocation of skills had been done, I carried on evaluating the percentage correct for each skill by candidate, and produced the same topic based table for skills, a snap-shot shown below:

	Manipulate	Identify	Simplify	Factorise	Evaluate
A	100	100	93	100	84
B	93	80	90	80	81
C	93	100	93	100	71
D	100	90	97	80	97
E	100	90	97	80	97
F	86	90	86	100	81
G	100	80	100	100	94
H	93	60	86	100	84
I	93	90	90	100	87
J	100	100	100	100	100
K	71	60	72	100	74
L	71	80	62	80	52

Figure 6.13: A snap-shot of the skills analysis by candidate, C1 June 08.

What this analysis was in part aiming to pick up was a handful of key skill areas for individual candidates to go away and work on with their teachers, but also some key topic and skill deficiencies across the centre that could be improved on with sharing of teacher resources and strategies, and also feedback and helpful tips from the exam board training and feedback training days. For candidate L, we would obviously recommend work on Simplifying and Evaluation questions.

	A	B	C	D	E	F	G	H	I	J	K	L	M
2	98 cands	**Draw**	**Eval**			**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	
3	Jun-07	91	89			90	93	97	90	92		91	
4	Rank	5	1			2	7	8	3	6		4	
5	Marks	6	29			20	23	9	9	31		18	
7	12 cands	**Draw**	**Eval**	**Factor**		**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	
8	Jan-08	89	88	72		79	83	85	92	88	91	81	
9	Rank	8	6	1		2	4	5	10	7	9	3	
10	Marks	9	24	3		20	28	8	3	26	9	12	
12	107 cands	**Draw**	**Eval**	**Factor**		**Iden**	**Manip**	**Prev**	**Show**	**Simp**		**Substit**	
13	Jun-08	88	88	96		90	95	92	95	91		94	
14	Rank	2	1	9		3	7	5	8	4		6	
15	Marks	6	29	5		10	12	7	5	27		20	
17	39 cands	**Draw**	**Eval**	**Factor**	**Frac**	**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	**Symbol**
18	Jan-09	85	86	92	99	79	89	86	77	92	95	92	83
19	Rank	4	6	10	12	2	7	5	1	9	11	8	3
20	Marks	11	19	2	4	15	17	5	5	16	2	16	3
22	82 cands	**Draw**	**Eval**			**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	
23	Jun-09	82	80			87	86	90	81	86	84	89	
24	Rank	7	9			3	5	1	8	4	6	2	
25	Marks	4	17			21	25	11	6	33	8	16	
27	36 cands	**Draw**	**Eval**			**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	
28	Jan-10	93	81			90	86			86	69	88	
29	Rank	7	2			6	4			3	1	5	
30	Marks	4	27			29	32			23	5	16	
32	374												
33	candidates												
35		**Draw**	**Eval**	**Factor**	**Frac**	**Iden**	**Manip**	**Prev**	**Show**	**Simp**	**Solve**	**Substit**	**Symbol**
36	Average	88	85	87	99	86	89	90	87	89	85	89	83
37	Difficulty rank (av)	7	3	5	12	4	8	11	6	9	2	10	1
38	Marks (av)	7	24	2	1	19	23	7	5	26	4	16	1
39													
40	Weighted Average	88	86	88	99	86	88	90	87	89	84	90	83

Table 6.14: Skills analysis across C1 papers for 374 candidates.

Conclusions from the whole set of results are equally important as topic based data:

- Use of symbols and remembering how to manipulate and simplify them is marginally the hardest skill to master, but occurs so infrequently (just in one paper) and is hardly worth many marks
- Solving / Evaluating expressions and equations are the next two hardest skills, again indicative of similar grasp of both skills. They occur very often with a combined total of 14 marks per paper on these skills.
- Simplifying / Manipulating occur very often and make up 25 marks on the paper – the largest share of any skill pairing, and both have the same success rating (89%), indicative of students' similar grasp of both skills.
- Identification / Substitution occur on every paper too, and account for 18 marks on each paper, again with similar success and levels of marks available.

What is really striking is that virtually every skill set scores between 80 and 90% success fairly consistently across the papers, across the years. The weakest years were January 08 (which is to be explained as all 12 (weaker than average) candidates were retaking the paper from the previous year), and January 09 (again mostly retake students from the previous summer sitting). Equally striking is the fact that the summer peak in June 08 (where we had more bright students sitting the paper for the first time) is not matched again by the results in June 09. After the June 08 paper, there seems to be a negative trend thenceforth. What has caused skills to decline slightly here and yet left topic success unaffected warrants further investigation, as shown in Figure 6.15:

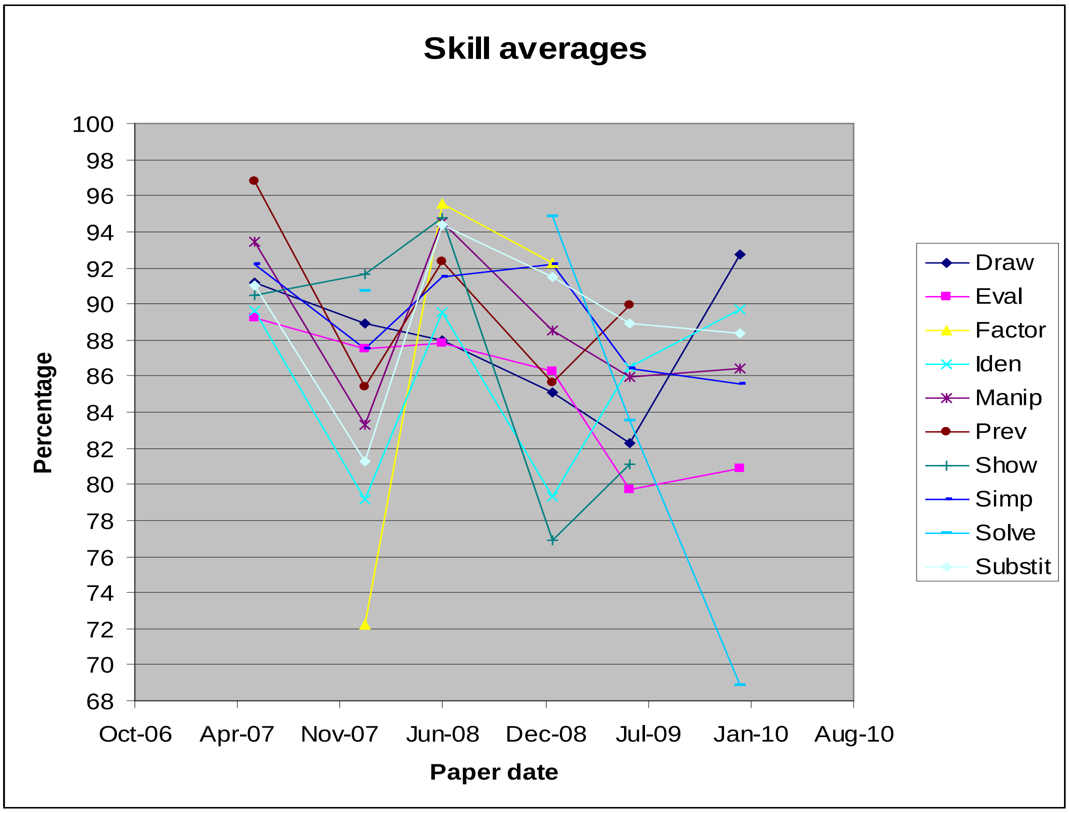


Figure 6.15: Skills analysis average percentage by paper.

6.5 Diagnosis of student errors

This research is also aiming to provide accurate diagnosis and categorisation of student errors. At present, the major usage of diagnosis of topic and skill deficiency is for candidates wishing to resit the exact same paper, although some tangible benefits can be derived by knowledge of specific weaknesses (if any are present) and the candidates don't wish to resit the paper. The first group of candidates I wish to study are a group of 12 students who first sat the paper in June 07. They all re-sat in January 08 (and were the only students entered for January 08), and then 8 of them re-sat the paper again in June 08. The topic analysis for the students makes very interesting viewing, in table 6.16:

Jun-07	**AP**	**Area**	**Differ**	**Form**	**Frac**	**Ind**	**Inequ**	**Int**	**Lines**	**Polyn**	**Pythag**	**Quadr**	**Sequ**	**Simult**	**Surd**	**Transf**	**Mark**
A	100		77	79		94	67	100	86	83		82	71	100		80	**64**
B	100		69	75		94	67	89	64	58		55	100	57		60	**56**
C	100		92	79		100	100	100	86	58		82	57	71		0	**62**
D	100		77	93		100	67	100	93	83		64	100	71		60	**66**
E	67		100	64		100	100	100	71	58		82	43	71		0	**58**
F	100		100	71		76	100	67	64	50		82	57	71		0	**54**
G	100		100	75		94	67	67	64	100		73	100	57		100	**62**
H	67		100	71		100	100	100	86	83		82	43	71		60	**63**
I	100		100	79		94	100	100	100	92		82	14	71		100	**65**
J	100		69	61		88	33	67	64	75		64	43	71		40	**52**
K	100		100	93		100	100	100	86	75		82	100	71		60	**68**

Jan-08	**AP**	**Area**	**Differ**	**Form**	**Frac**	**Ind**	**Inequ**	**Int**	**Lines**	**Polyn**	**Pythag**	**Quadr**	**Sequ**	**Simult**	**Surd**	**Transf**	**Mark**
A	100		86	95	100	89	71	80	73	69	100	90	100		100	43	**64**
B	43		43	68	75	89	29	90	64	85	67	45	75		100	71	**50**
C	100		100	95	75	95	100	100	100	92	100	100	100		100	100	**73**
D	57		100	77	75	95	71	100	100	69	100	90	75		100	43	**63**
E	57		86	73	75	95	57	100	100	85	100	65	75		60	86	**60**
F	57		79	77	75	58	71	70	91	85	67	85	63		80	71	**56**
G	57		100	86	50	84	86	100	100	100	100	80	100		20	100	**65**
H	86		100	95	50	89	71	100	100	100	100	90	100		100	100	**70**
I	57		100	86	75	95	100	100	100	100	100	100	100		100	100	**71**
J	57		93	68	75	63	29	40	73	100	100	60	75		40	100	**52**
K	100		100	95	100	100	57	100	91	100	100	85	100		100	100	**71**

Jun-08	**AP**	**Area**	**Differ**	**Form**	**Frac**	**Ind**	**Inequ**	**Int**	**Lines**	**Polyn**	**Pythag**	**Quadr**	**Sequ**	**Simult**	**Surd**	**Transf**	**Mark**
A	100	100	80	100		56	100	33	88	100	100	100	100			80	**65**
B	100	75	67	100		100	100	89	75	100	100	100	100			100	**68**
E	100	100	93	100		56	100	33	96	100	100	100	100			60	**66**
F	100	0	73	100		100	100	89	63	75	100	100	100			60	**61**
G	80	0	100	87		100	100	100	79	92	100	100	100			100	**68**
H	100	100	100	100		100	80	100	100	100	100	94	100			100	**74**
J	100	0	93	100		56	80	33	75	75	100	82	100			100	**60**
K	100	100	100	100		100	100	100	100	100	100	100	100			100	**75**

Figure 6.16: Results topic analysis for 12 of group A retakes.

As is plainly clear from the retake analysis:

- Of those who thought they had mastered Arithmetic Progressions (AP) and Inequalities on the first paper, it clearly tripped them up badly 2nd time round, but the improvement on the 3rd paper is remarkable having received such "feedback" in the form of searching questions in their 2nd attempt.
- The scores on transformations for each group improve from 51% average to 83% average 2nd time round to 88% average third time, suggesting how much work had gone in to improving and learning the material on this topic between retakes. The chart of averages by topic is below in figure 6.17:
- Quadratics, Sequences, Formula and Polynomials all improve similarly to transformations, suggesting vast improvements in revision and practice.
- However, Differentiation stays almost exactly the same, Indices drops in success by over 10%, and Integration by nearly 20% by the 3rd attempt. Straight lines improve very well before dropping slightly on the 3rd attempt.

	AP	Area	Differ	Form	Frac	Ind	Inequ	Int	Lines	Polyn	Pythag	Quadr	Sequ	Transf
Jun-07	93.939		89.51	76.299		94.652	81.818	89.899	78.571	74.242		75.207	66.234	50.909
Jan-08	70.13		89.61	83.471	75	86.603	67.532	89.091	90.083	89.51	93.939	80.909	87.5	83.117
Jun-08	98	59	88	98		83	95	72	84	93	100	97	100	88

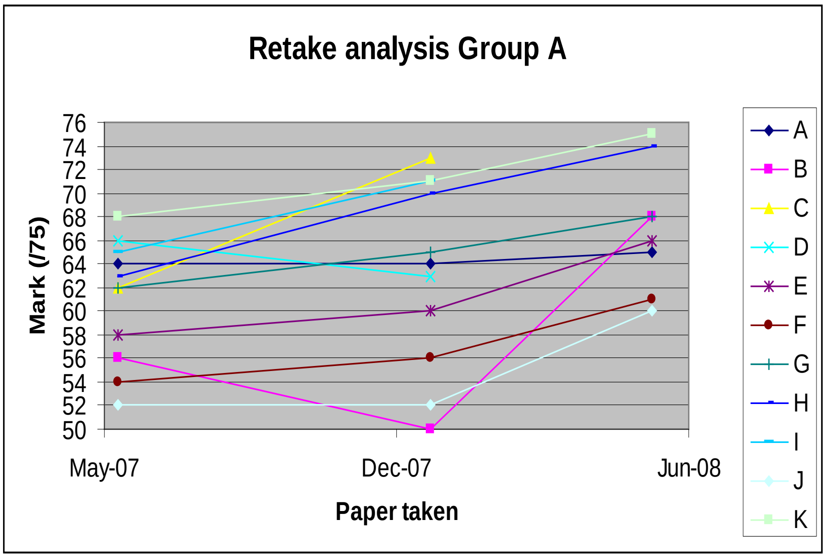
Figure 6.17: Changes by topic average for each retake.

There are clearly some very important conclusions to be drawn here, and the major revision topics of Indices and Integration have been neglected between retakes, despite how much they are built on in future core maths papers. Some topics, like Arithmetic progressions will be easily forgotten if not revised again, as they are more stand-alone in nature, so will need additional time and practice devoted to them.

What is remarkable is how little changes are made from the 1st attempt to the 2nd attempt. Clearly very little additional work has gone on to improve on the paper, but this changes remarkably by the 3rd attempt. Candidates gained only 2 more marks 2nd time round, but had improved by 7.5 marks on average (13%) by the 3rd attempt on the paper. Interestingly four candidates failed to improve at all (out of the 12) after the 2nd attempt, and one candidate managed to improve by 32% over the 3 attempts, as shown below in figures 6.18 and 6.19.

Figure 6.18: Changes to scores after 1 or 2 retakes.

	Jun-07	Jan-08	Jun-08
A	64	64	65
B	56	50	68
C	62	73	
D	66	63	
E	58	60	66
F	54	56	61
G	62	65	68
H	63	70	74
I	65	71	
J	52	52	60
K	68	71	75



	$2^{nd} - 1^{st}$	% change
A	0	0.00
B	-6	-10.71
C	11	17.74
D	-3	-4.55
E	2	3.45
F	2	3.70
G	3	4.84
H	7	11.11
I	6	9.23
J	0	0.00
K	3	4.41
Av	2.27	3.57

	3rd - 1st	% change	3rd - 2nd	% change
A	1	1.56	1	1.56
B	12	21.43	18	32.14
E	8	13.79	6	10.34
F	7	12.96	5	9.26
G	6	9.68	3	4.84
H	11	17.46	4	6.35
J	8	15.38	8	15.38
K	7	10.29	4	5.88
Av	7.50	12.82	6.13	10.72

Table 6.19: Retake analysis from 3rd to 2nd and 2nd to 1st papers.

6.6 Conclusions and future direction

We can equally easily build up a difficult and challenging paper (like in the first chapter), based on those topics that ranked 1^{st} or 2^{nd} by difficulty in any year as follows:

June 07 - Transformation and Simultaneous Equations
January 08 – Inequality and Arithmetic Progression
June 08 – Area and Integration
January 09 – Straight Lines and Quadratic Equations
June 09 – Differentiation, Surds and Indices
January 10 – Indices and Integration

Such a paper would be a very good challenge to students, and I have put it together in appendix 3 as an exercise in writing challenging preparation material based on what previous students have found hardest. It covers virtually the whole syllabus, and would suffice as an excellent practice exam to stretch talented students. I have generated a reusable version of this paper in Excel format (as in chapter 5), and a screenshot of such a C1 dynamic paper in appendix 4. Again, as in chapter 5, this process is time-consuming, fiddly and the format is difficult to really make user-friendly. The present purpose of this chapter is to elicit useful error analysis and student feedback from ResultsPlus files. Such reusable papers lead in to the following chapter on analysis of and writing future CAA materials. CAA test data will allow much more refined topic analysis by errors and also allow tailor made tests for candidates / cohorts to be made to help improve on specifics. Where this analysis is useful for teaching staff, is that they readily use non-CAA methods in the classroom and tailor make lessons and exam preparation based on such specific topic areas and the nature of previous exam questions by repetition.

The next stage of development here is the realisation that to be successful at retaking exam papers, we need further accurate diagnosis of errors and skills / topic improvements. We also need tailor-made questions easily accessible to students so that they can self-motivate to prepare for the retakes, especially when class time is often devoted to new courses rather than retakes. This mode of analysis, while very informative to the teacher, and quite informative to the student, really doesn’t give the student many further practice materials or skill or topic specific questions that they can go away and practise repeatedly. The format of ResultsPlus is also not as informative as we would hope in categorising student errors as our pilot paper-based (and onscreen) test was. CAA can answer all of these requests.

Chapter 7 – Analysis of a CAA test on Differentiation

7.1 CAA Tests

Brunel University, along with almost every other tertiary institution, has recognised the benefits of running diagnostic testing for 1st year undergraduate students, as well as foundation year students. They have been using formative and diagnostic computer aided assessment (CAA) for several years now, primarily for mathematics students, and more recently for students on courses that required a high degree of mathematical skills to underpin future learning (Psychology, Economics, IT). For Economics students, understanding the foundations of Calculus (in particular differentiation) is key to understanding future economic courses.

Given our major question, “do students make the same mistakes on paper as online?”, touched upon in the development of our online pilot test (in chapter 5) from its paper-based version, then we are motivated to ask similar questions using CAA to help answer this. We developed a taxonomy of student errors in chapter 3 and 4 equally based on differentiation, so it would be prudent to explore the use of CAA in diagnosing and categorising student errors also through differentiation. I am also keen to see CAA develop tests spanning a range of skills, not just for diagnosis purposes, but also to develop accurate and reusable formative assessments.

A sample of differentiation questions was tested on Economics first year students (course code: EC1005) at Brunel University during 2008-2009 and 2009-2010, **Greenhow (2010)**. The Economics students sitting level 1 (a compulsory course in the 1st semester) are largely expected to achieve 300 UCAS points for entry to the course, and level 1 students must have at least AS mathematics course (thus incorporating many foundation skills in calculus). About 80% of 2008-2009 students have achieved A Level mathematics, in roughly equal proportions of each grade A to E, **Greenhow (2009)**.

The objectives of these tests were as follows:

- To diagnose basic calculus deficiencies at an early stage, and allow for correction
- To homogenise the mathematical levels of the cohort (who had a range of backgrounds in maths from GCSE through to A-Level)
- To understand basic differentiation rules and apply them to differentiate polynomials and algebraic functions (lower level skills)
- To understand the product rule and apply it to polynomials and exponentials (mid level skills)
- To understand the chain rule and apply it to harder questions on binomial functions (hard level skills)

In these tests, the practical application of calculus to Economics was not being tested, and this link and usage is expected to have been introduced to students at an early stage to encourage motivation for learning the harder rules. As set out in the aims and objectives above, I will also subdivide the test questions into five distinct skill levels each with a difficulty tag, and aim to compare performances across the skill levels with the facilities generated in the tests, as detailed in Table 7.1. The difficulty tags (cognitive groupings) are used primarily to cluster the skills together based on similarity of scores in table 7.2

Skill descriptor	Difficulty tag
Differentiation\Polynomials	Lower
Differentiation\Product rule\Polynomials	Middle
Differentiation\Algebraic functions	Lower
Differentiation\Chain rule\Binomials	Higher
Differentiation\Product rule\Exponentials	Middle

Table 7.1: Skill descriptors for questions used in EC1005 assessment.

The 35 differentiation questions were offered as part of formative objective tests for these students, and they were allowed up to 7 attempts to complete the tests. The test comprised mostly Multiple-Choice (MCQ) questions (29), with 3 Numerical Input (NI) questions, 1 Hot Line (HL) question (where an erroneous proof is given and a "hot" error line is needed to be chosen) and 1 Responsive Numerical Input (RNI) question. Results from both 2008-2009 and 2009-2010 are summarised in figure 7.2 by facility value (given as a %):

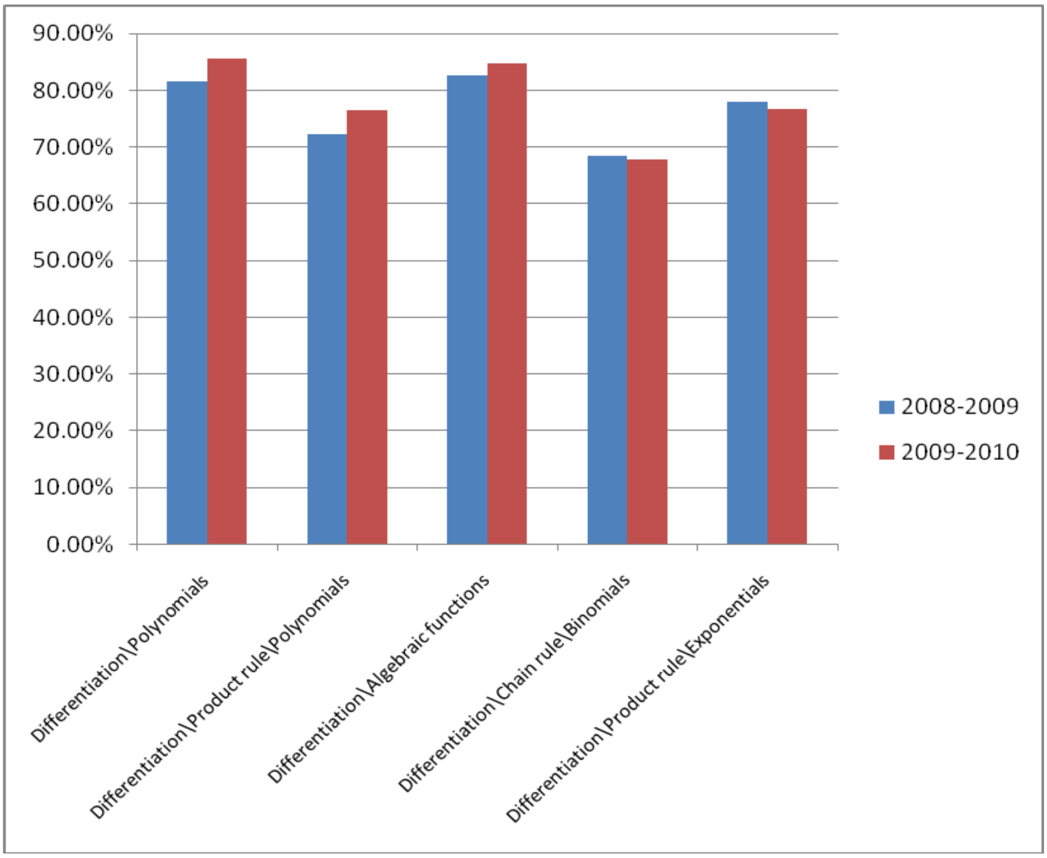


Figure 7.2 Facility by skill area for 2008-2009 and 2009-2010 students on EC1005 courses

I personally recognise the difficulties students have at a higher cognitive level in grasping and using the chain rule on binomial functions at A-Level, and if 20% of students on EC1005 have never seen the last four skill areas, and many of those will grades C, D and E on A Level maths have never grasped these skill areas, then we already have an area to explore in more depth by evidence through the results generated by these two cohorts. If we can break down the skills needed to complete a differentiation question using the chain rule for binomial functions, (a multi-structural problem), we can hopefully identify the uni-structural mistakes students make, and go back a few levels to fewer-structured problems on similar skills to work out where the underlying domain of knowledge resides, and where the remedial work needs to begin.

Within these tests, 854 performances were returned in 2008-2009 (at an average of 2.44 tests per student), and 643 tests were returned in 2009-2010 (at an average of 1.84 tests per student). The 35 questions (of which a full list by skill level is included in Appendix 5) were offered to students in 2008-2009, and of these questions, only one was not present in the database offered to 2009-2010, and is thus only included for reference and for the initial Mann-Whitney hypothesis test, but is rejected for the basis of comparison. The rejected question (present in 2008-2009, not in 2009-2010):

Find $\frac{d}{dx}\left(ax^m + bx^p\right)^n$, whereby both “a” and “n” are positive integers (i.e. question realisations given with distracters calculated on the “a” and “n” random parameters.

Within the analysis generated by results files from the assessment tests, question descriptors appear as the following (a, b are fractions):

sqrt((x^a)^b); MC	Differentiation\Algebraic functions

This refers to: Multiple choice questions on Differentiation of Algebraic Functions.

In mathematical form, it appears onscreen as: Find: $\frac{d}{dx}\left(\sqrt{\left(x^a\right)^b}\right)$

7.2: Comparison of student performances across skill levels

A Mann-Whitney test is used to compare the averages of these three cognitive grouping in section 7.2, to demonstrate whether there are inherent differences in the skill levels required for each cluster of topics.

To carefully analyse student performances, the reports generated by the software produce data for each question for all completed attempts. Of 3593 attempts in 2008-2009, 3577 were completed with an answer (99.56%). In 2009-2010, 2516 questions were attempted, and 2494 were answered (99.13%). For reasons unknown to us, students may have exited the tests so leaving this tiny percentage unanswered, and thus we must ignore this from the analysis. The full spreadsheet listing results for both cohorts can be found on the **Differentiation** tab of the **EC1005_CAA_question_stats.xls**. The summary statistics for both groups (including weighted mean (facility) and standard deviations) are shown below:

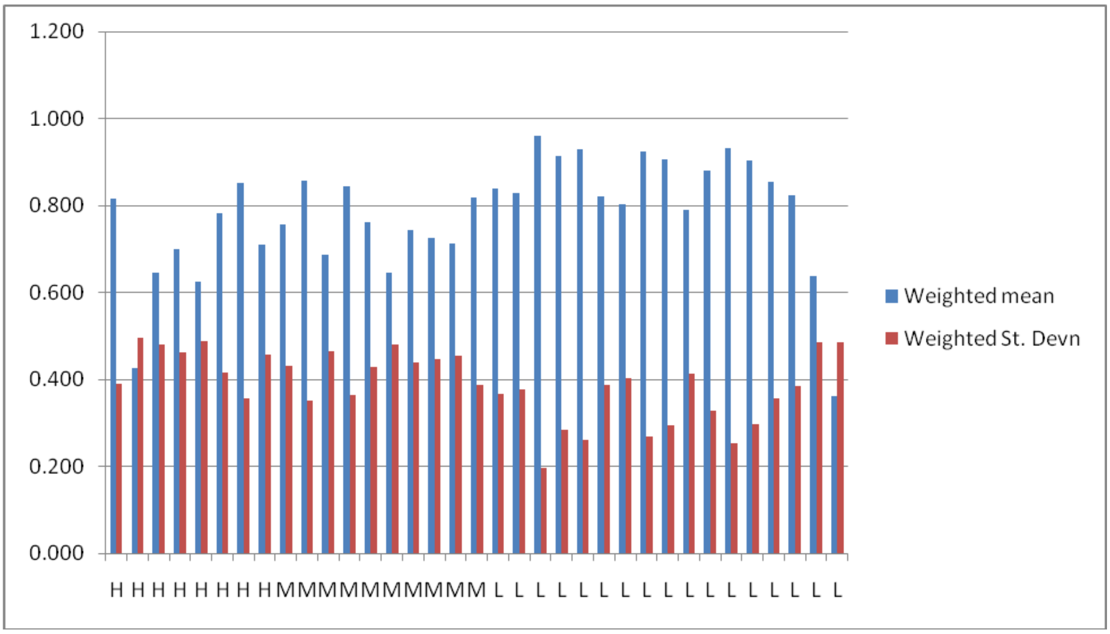


Figure 7.3: Weighted mean and standard deviation by cognitive area for EC1005.

Question descriptor	Level	Facility	Weighted St. Devn	Rank Means
sqrt((x^a)^b); MC	L	0.840	0.367	13
sqrt(x^a); MC	L	0.829	0.377	14
(ax^b+c)^n,x,a,b,c); MC	H	0.817	0.389	18
(ax^m+bx^p)^n; a,n +ve; MC	H	0.427	0.497	34
(ax+b)^3; HL	H	0.647	0.481	30
(ax+b)^n; a,b +ve/-ve,n+ve; NI	H	0.700	0.462	28
(ax+b)^n; a,n +ve/-ve; RNI	H	0.625	0.488	33
(ax+b)^n; a,n +ve; MC	H	0.783	0.415	21
(ax+bx^3)^3; a,b +ve; MC	H	0.852	0.357	11
(b+cx)^3 ; b,c +ve; MC	H	0.709	0.457	27
diff ax^b; a,b +ve; MC	L	0.961	0.197	1
diff cubic version 1; MC	L	0.914	0.285	5
diff cubic version 2; MC	L	0.929	0.261	3
diff cubic; +ve coeffs; evaluate at x=1/a; MC	L	0.820	0.388	16
diff cubic; +ve coeffs; evaluate at x=a; MC	L	0.802	0.402	19
diff polynomial n=4...12 version 1; MC	L	0.924	0.268	4
diff polynomial n=4...12; MC	L	0.906	0.294	6
diff polynomial n=4...6; +ve coeffs; eval at x=1/a; MC	L	0.789	0.412	20
diff polynomial n=4...6; +ve coeffs; eval at x=a; MC	L	0.879	0.329	8
diff quadratic version 1; MC	L	0.932	0.254	2
diff quadratic version 2; MC	L	0.904	0.298	7
diff quadratic; +ve coeffs; evaluate at x=1/a; MC	L	0.855	0.356	10
diff quadratic; +ve coeffs; evaluate at x=a; MC	L	0.824	0.384	15
Min of f(x)=Ax^2+Bx+C, A, B, C +ve; NI	L	0.639	0.487	32
x+y given. Find y for TP of xy^2. x,y +ve; NI	L	0.361	0.486	35
diff(cubic* Exp(cubic); MC	M	0.756	0.432	23
diff(quadratic* Exp(ax); MC	M	0.856	0.352	9
diff(quadratic* Exp(cubic); MC	M	0.687	0.466	29
diff(quadratic* Exp(linear); MC	M	0.844	0.365	12
diff(quartic* Exp(quadratic); MC	M	0.762	0.428	22
diff cubic*cubic; MC	M	0.645	0.480	31
diff quadratic*cubic; MC	M	0.743	0.439	24
diff quadratic*quartic; MC	M	0.726	0.448	25
diff quartic*quartic; MC	M	0.713	0.454	26
diff(a+bx^2+cx^3)(dx+ex^3+fx^5); MC	M	0.818	0.387	17

Table 7.4: Weighted mean and standard deviation for both cohorts.

The weighted mean (facility value) of the questions clearly rises (0.695, 0.755, 0.83) as we progress from High to Low cognitive levels, with a couple of notable exceptions (Question 2 of High level and the last two questions of Low level). This Question 2 has previously been identified as not appearing on cohort 2009-2010:

Find $\frac{d}{dx}\left(ax^m + bx^p\right)^n$, whereby both "a" and "n" are >0, and as such is discounted from the overall conclusions, yet included here for reference.

The weighted standard deviation falls as we progress from High to Low cognitive levels (0.443, 0.425, 0.344), indicative of much closer scoring on these questions. Overall, the weighted mean score of 76.6 % in 2008-2009 increased slightly to 78.3% in 2009-2010, but this is not statistically significant over the number of tests carried out.

However, a Mann-Whitney test on the differences between the averages of these cognitive levels can be carried out. If we suppose, as a null hypothesis, that there is a significant distinction between performances on questions requiring higher cognitive skills than lower ones, we can test for this difference by testing the Low and Middle skills, and then testing the Middle and Higher skills questions in a separate test. If we draw conclusions that there is significant difference on either of these comparative 2-sample tests, it stands to reason that there will be statistically significant differences between the lower and higher cognitive skill groups.

Thus a null hypothesis, H_0 would state: There is no significant difference between performances on questions covering the lower cognitive skills to those covering middle cognitive skills. As opposed to an alternative hypothesis, H_1: There will be a significant difference between performance on questions covering the lower cognitive skills to those covering middle cognitive skills. We assume data to be taken uniformly from both years, thus this test does not test between the years (i.e. no significant temporal changes).

This is a non-directional (2-tailed) hypothesis test, and as a first assumption, I will look to carry out the test at a 5% significance level, so that we have a reasonably good degree of confidence in our conclusions. We use a Mann-Whitney test, because is a non-parametric test for assessing whether two independent samples of observations come from the same distribution. In this first case, we want to test whether or not the samples of lower cognitive questions are from the same distribution as the middle cognitive ones. Then we can test whether the samples of questions from middle cognitive levels are from the same distribution as those from higher cognitive levels.

Following **Mann & Whitney (1947)**, we group observations (in this case the mean / facility score per question) into the two groups and rank them against the whole set of 27 questions. The groups contain 10, 17 questions respectively (for Middle and Lower cognitive level skills). Thus we can't consider this a paired test (and it is worth noting that we are not confident the samples came from the same population). At the non-directional 5% level (i.e. the probability that we incorrectly reject the null hypothesis is 5% or less), we require U (test statistic) to be 45 or more. The respective sum of ranks for group 1 and 2: R1, R2 respectively, with N1 and N2: number in each group respectively. Each of our two tests will require the calculation of "U" values for each group, and the summary statistics are given in table 7.5 below.

Mann-Whitney		H + M	M + L
R1		88	190
R2		83	188
N1		8	10
N2		10	17
U1		52	135
U2		28	35
Crit values		17	45
Lower U		28	35
Test result		Accept Ho	Reject Ho
U1 + U2	check	80	170
N1 x N2	check	80	170

Table 7.5: Mann-Whitney U test values for Groups Higher + Middle (H + M) and Middle + Lower (M + L) skills of EC1005 across 2008-2010 inclusive.

Given that N (N1 and N2) is less than 20 for each sample, we won't have justification for using a normal approximation to the distribution, but the results are very clear nonetheless. Taking the lower U values (28 and 35 respectively), we see this is below the critical value in the Middle and Lower skill question comparison, but above it in the Higher and Middle skill comparison (as given in table 7.6 below), courtesy of: http://math.usask.ca/~laverty/S245/Tables/wmw.pdf

So we can reject H_0 in favour of H_1, that there is a statistical difference between the middle and lower skills samples, namely that there is a significant difference between performance on middle cognitive questions than lower cognitive ones. Even if the significance level of the test was reduced to 0.01 (non-directional), we still have a critical value for U of 34, given n1 = 10, n2 = 17, so we would almost draw the same conclusion, only with a more stringent test and confidence interval.

However, the test failed to reject Ho for the comparison between middle and higher skill questions, namely that there is insufficient evidence to show that there is a difference in performance on middle skill questions as on higher skill questions. This could indicate that the lower to middle skill overlap should be the major area for testing and thus differentiating between the students. More work is also still needed for students preparing for test questions at a higher cognitive level, but this test may suggest that weaker students who can do lower skill questions are likely to be equally stumped by middle and higher skill questions.

Table A5.07: Critical Values for the Wilcoxon/Mann-Whit

Nondirectional α=.05 (Directional α=.025)

	n_2																
n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45

Table 7.6: Critical values for the Mann-Whitney U test at the 5% non-directional level.

7.3: Comparison of student performances and analysing differences between tests

There is shown to be a slight improvement in average score (76.6% to 78.3%) from 2008-2009 to 2009-2010 and such a difference is again shown up when we look at the difference between the facility values on each question. The correlation between the two sets of data is measured as 0.853 (Spearman) and 0.854 (Pearson), indicating a very good agreement with question by question scores and limited major changes either on single questions or across cognitive skill areas. One significant difference between the 08/09 and 09/10 cohort is that the later one have had exposure to Mathletics throughout the academic year, and all other factors appear to be the same, so we should be tempted to conclude that the computer-aided learning has enhanced the scores from one year to another. The average facility increase per question is 0.03 – a 3% increase across questions, which is fairly well mirrored across skill levels. Lower skill questions increase by 3.2%, middle skill questions by 3.8%, and higher skill questions by 2.6%. Given that the improvement in marks across the skill levels are very similar, suggests that the whole cohort has improved from one year to the next.

However what catches the eye are the questions where the difference is of the order of 0.1 (where we know that there has been a major change, as the question facility has increased by 10% or more). In this case, I can highlight four such questions (in yellow) in table 7.7.

The other major spin-off of these results in table 7.7 is that the questions that have decreased by the most in facility value could also be included in a new test for future groups that really allows good students to be stretched and allows those who haven't done enough practise to be found out if they have been reliant on "easier" questions coming up randomly in the test. Such a test is detailed in Appendix 6.

Question descriptor	Difference of means
sqrt((x^a)^b); MC	0.04
sqrt(x^a); MC	-0.002
(ax^b+c)^n,x,a,b,c); MC	0
(ax+b)^3; HL	-0.03
(ax+b)^n; a,b +ve/-ve,n+ve; NI	0.083
(ax+b)^n; a,n +ve/-ve; RNI	-0.033
(ax+b)^n; a,n +ve; MC	0.016
(ax+bx^3)^3; a,b +ve; MC	0.198
(b+cx)^3 ; b,c +ve; MC	-0.053
diff ax^b; a,b +ve; MC	0.024
diff cubic version 1; MC	0.01
diff cubic version 2; MC	0.053
diff cubic; +ve coeffs; evaluate at x=1/a; MC	0.044
diff cubic; +ve coeffs; evaluate at x=a; MC	0.004
diff polynomial n=4...12 version 1; MC	0.005
diff polynomial n=4...12; MC	0.009
diff polynomial n=4...6; +ve coeffs; evaluate at x=1/a; MC	-0.023
diff polynomial n=4...6; +ve coeffs; evaluate at x=a; MC	0.114
diff quadratic version 1; MC	-0.037
diff quadratic version 2; MC	0.045
diff quadratic; +ve coeffs; evaluate at x=1/a; MC	0.05
diff quadratic; +ve coeffs; evaluate at x=a; MC	0.038
Min of f(x)=Ax^2+Bx+C, A, B, C +ve; NI	0.144
x+y given. Find y for TP of xy^2. x,y +ve; NI	0.023
diff(cubic* Exp(cubic); MC	0.071
diff(quadratic* Exp(ax); MC	-0.021
diff(quadratic* Exp(cubic); MC	0.064
diff(quadratic* Exp(linear); MC	-0.062
diff(quartic* Exp(quadratic); MC	-0.025
diff cubic*cubic; MC	0.093
diff quadratic*cubic; MC	0.149
diff quadratic*quartic; MC	0.008
diff quartic*quartic; MC	0.087
diff(a+bx^2+cx^3)(dx+ex^3+fx^5); MC	0.012

Table 7.7: Results of differences by question 2008-2009 and 2009-2010 of EC1005.

Clearly much work has been done between the year-groups (whether this has been noticeably different in lectures, practice materials or indeed in the background of the cohorts) that the following questions have been significantly better answered:

Question descriptor	Difference of means
(ax+bx^3)^3; a,b +ve; MC	0.198
diff polynomial n=4...6; +ve coeffs; evaluate at x=a; MC	0.114
Min of f(x)=Ax^2+Bx+C, A, B, C +ve; NI	0.144
diff quadratic*cubic; MC	0.149

Table 7.8: Positive improvements in question response from EC1005 2008-2010

The first question of these questions is the only one that has been designated a high level skill, and the next two questions low level skills with the last question a middle level skill. Clearly much work has been done over the year to improve on certain aspects of the understanding of the chain rule, and I could hypothesis that the lecturer has specialised in a very similar example question to teach students these skills, so such a question was this year better answered. The high level question increased its facility value from 0.763 to 0.961, again indicating help was given as this facility value is almost anomalous within the higher cognitive level question group. Curiously and anecdotally, a very similar question was asked in the cohort EE1083 (first year electrical and computer engineering students at Brunel), who sat the same question: (ax+bx^3)^3; a,b +ve; MC , with a and b given as 1. This question yielded the highest discrimination of any question on their papers on calculus, which would suggest that this type of question has caused problems in the past and has thus been looked at between students across courses over the last year.

7.4 Discrimination differences and high-lights between the tests

The higher the discrimination / correlation, the better a question is in assessing the skills required to be tested.

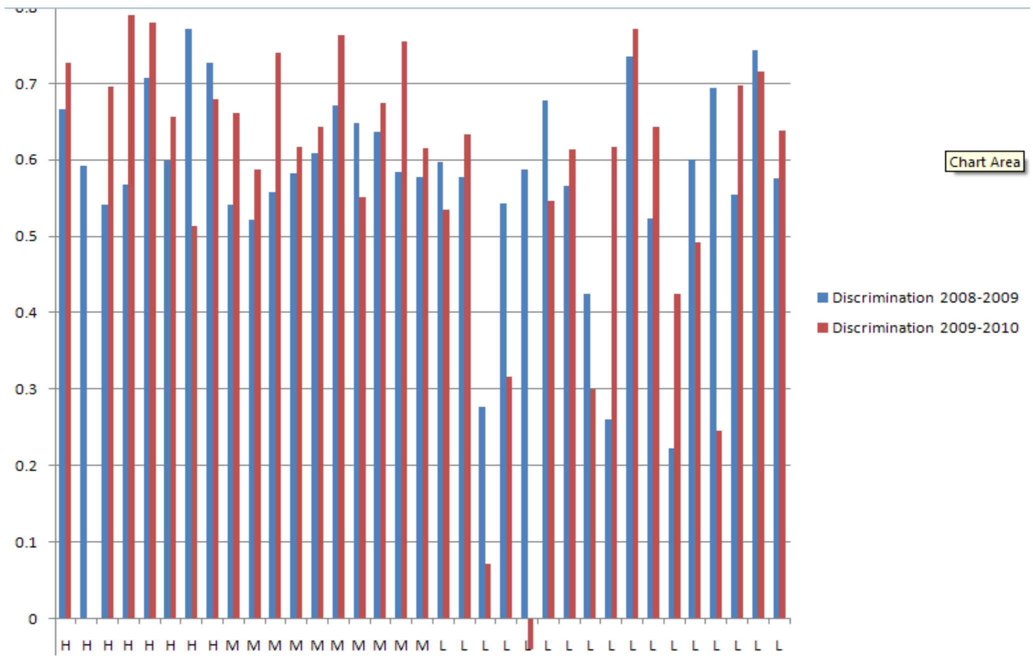


Figure 7.9: Discrimination by skill level compared by year.

In figure 7.9 (above), we have a very clear trend from high level questions to low level questions (0.65, 0.59, 0.54 in 2008-2009 and 0.69, 0.66, 0.48 in 2009-2010) whereby, in this particular case, the higher the skill level, the higher the discrimination index. The one negative discrimination result, could indicate that weaker students very slightly tending to do better on the question that the stronger students. This is not unknown as a poorly designed question could throw strong students off but not the weaker ones. One observation is that low level questions have dropped by 0.056 in discrimination index from 2008-2009 to 2009-2010, below 0.5, indicating possibly that lower level questions have become weaker at differentiating between students than harder or middle cognitive level questions should be used to assess such cohorts. One hypothesis is that students have spent much time working on these easier cognitive level questions on Mathletics and mastered the basic skills without spending as much time on the harder level questions. Thus good students can still make errors on low skill questions and weak students don’t necessarily get the low skill questions correct either (symptomatic of strategic guesswork or pre-structural responses in some cases). It is also worth pointing out that the average discrimination didn’t change from one year to the next.

Table 7.10 below: Results of discrimination year by year and the associated differences:

Question descriptor	Level	2008-2009	2009-2010	Difference
(ax^b+c)^n,x,a,b,c); MC	H	0.667	0.727	0.060
(ax+b)^3; HL	H	0.541	0.696	0.155
(ax+b)^n; a,b +ve/-ve,n+ve; NI	H	0.568	0.790	0.222
(ax+b)^n; a,n +ve/-ve; RNI	H	0.707	0.779	0.072
(ax+b)^n; a,n +ve; MC	H	0.599	0.656	0.057
(ax+bx^3)^3; a,b +ve; MC	H	0.771	0.513	-0.258
(b+cx)^3 ; b,c +ve; MC	H	0.727	0.679	-0.048
diff(cubic* Exp(cubic); MC	M	0.542	0.662	0.120
diff(quadratic* Exp(ax); MC	M	0.521	0.587	0.066
diff(quadratic* Exp(cubic); MC	M	0.557	0.741	0.184
diff(quadratic* Exp(linear); MC	M	0.583	0.617	0.034
diff(quartic* Exp(quadratic); MC	M	0.608	0.643	0.035
diff cubic*cubic; MC	M	0.672	0.763	0.091
diff quadratic*cubic; MC	M	0.648	0.551	-0.097
diff quadratic*quartic; MC	M	0.637	0.675	0.038
diff quartic*quartic; MC	M	0.584	0.755	0.171
diff(a+bx^2+cx^3)(dx+ex^3+fx^5); MC	M	0.577	0.615	0.038
sqrt((x^a)^b); MC	L	0.598	0.534	-0.064
sqrt(x^a); MC	L	0.577	0.633	0.056
diff ax^b; a,b +ve; MC	L	0.276	0.072	-0.204
diff cubic version 1; MC	L	0.543	0.316	-0.227
diff cubic version 2; MC	L	0.588	-0.040	-0.628
diff cubic; +ve coeffs; evaluate at x=1/a; MC	L	0.678	0.546	-0.132
diff cubic; +ve coeffs; evaluate at x=a; MC	L	0.566	0.613	0.047
diff polynomial n=4...12 version 1; MC	L	0.425	0.299	-0.126
diff polynomial n=4...12; MC	L	0.26	0.617	0.357
diff polynomial n=4...6; +ve coeffs; evaluate at x=1/a; MC	L	0.736	0.771	0.035
diff polynomial n=4...6; +ve coeffs; evaluate at x=a; MC	L	0.524	0.644	0.120
diff quadratic version 1; MC	L	0.223	0.425	0.202
diff quadratic version 2; MC	L	0.601	0.492	-0.109
diff quadratic; +ve coeffs; evaluate at x=1/a; MC	L	0.694	0.245	-0.449
diff quadratic; +ve coeffs; evaluate at x=a; MC	L	0.554	0.697	0.143
Min of f(x)=Ax^2+Bx+C, A, B, C +ve; NI	L	0.744	0.716	-0.028
x+y given. Find y for TP of xy^2. x,y +ve; NI	L	0.576	0.638	0.062

The following question:

(ax+bx^3)^3; a,b +ve; MC	H	0.771	0.513	-0.258

provided much evidence of increase in performance and facility from 2008-2009 to 2009-2010, was the question in 2008-2009 with highest discrimination for the foundations of IT: EE1083, yet its discrimination has dropped by more than 25% from one year to the next, backing up my hypothesis that specific target coaching for this exact question had gone on, as its ability to discriminate has dropped it to the bottom 8 questions in usefulness on this scale in 2009-2010. Looking overall, the higher level questions have increased in discrimination by 0.037, the middle skill questions have increased by 0.068, while the lower skill questions have decreased by 0.056 in discrimination. Thus we can very readily draw the conclusions that from one year to the next, the middle and (to a lesser extent) higher cognitive level questions are proving the better discriminators, while the lower cognitive questions are much better answered and are a poor guide to subsequent performance on harder skill questions. All of the low discrimination values appear in the low cognitive level questions. When we compare the facility index with discrimination score on the overall set of questions, we also see an interesting result:

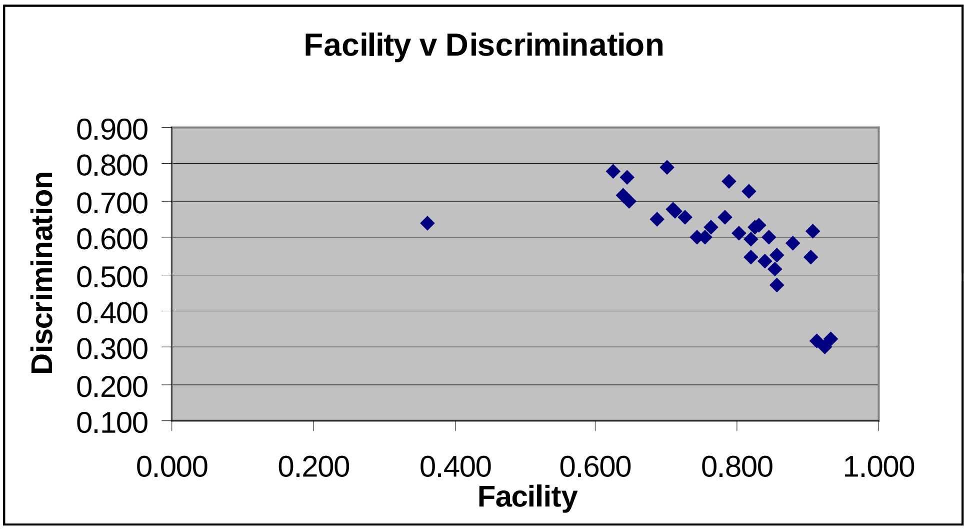


Figure 7.11: Facility and Discrimination scores for the EC1005 cohort from 2008-2010 inclusive.

Not only does this test represent a very good test, as all of the questions have facility values over 0.3, and all bar 2 of them over 0.6, but the discrimination index is above 0.5 for all questions bar 4. However, **Baruah (2007)**'s classification of difficulties for tests would classify 27 of our 34 questions as "easy". This classification would render the level descriptors meaningless and the relative facility values equally so, thus I propose judging the questions on their discrimination. After all, the graph above shows a very good negative correlation between facility and discrimination of -0.6254 (Pearson). We could question the internal consistency of such a test given this correlation index, but it makes sense to look at the harder questions. As the facility decreases so the discrimination increases, and thus the harder questions clearly discriminate between the students, something that our Mann-Whitney U test equally told us. A similar conclusion can be drawn by subdividing the questions into their levels.

As seen in Figure 7.12 below, facility is negatively skewed for this data (the bulk of values between 0.6 and 1.0), whereas discrimination is much more normally distributed. For robust comparisons, we would like Facility values to be normally distributed (0.55 – 0.65) and Discrimination very much skewed (near to unity).

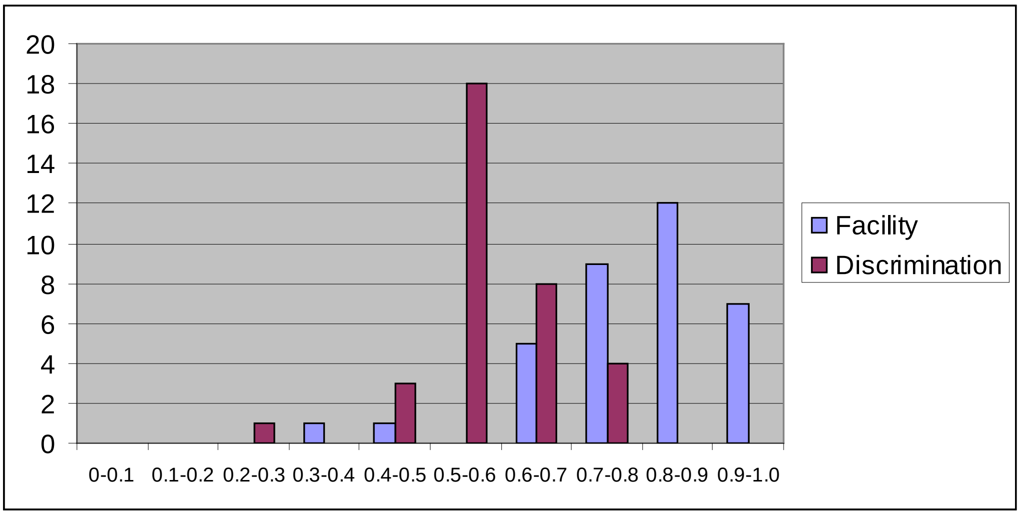


Figure 7.12: Facility and Discrimination scores across EC1005 between 2008-2010 inclusive.

What remains to be understood therefore, are the specific errors students make, in light of these easier questions, to ascertain where exactly the main understanding ceases and guesswork creeps in, or indeed up to which point the students' scores are really representative.

7.5 Question analysis and Specific student errors (mal-rules)

Looking through the answer files for the 35 differentiation questions yields some important clues as to why some questions prove harder than others, why some are better differentiators and why some can be very informative as to the specific student error (particularly if it is replicated across the cohort in any significant quantity). There are sufficiently few questions that are not MCQ (the RNI, NI and HL questions) to be unable to draw conclusions from them, as they a) only yield facility value, and b) the facility values lies in the middle of the distribution for their level, so are not significantly harder or easier than other question types.

I start by looking at a selection of questions from the lower cognitive level and then progress through to some medium cognitive level as there is a statistical difference between performances at these two levels (Mann-Whitney U test). At the lower cognitive levels (differentiation of algebraic functions), there were two questions set, both of which were multiple choice, and yielded facilities of 0.823 and 0.83, with discrimination indexes of 0.598 and 0.577 respectively. So both were easy questions that yielded a high number of correct answers but which told us relatively little (lower discrimination) about the likely overall performance given success on these questions. The harder of the two questions is given in figure 7.13:

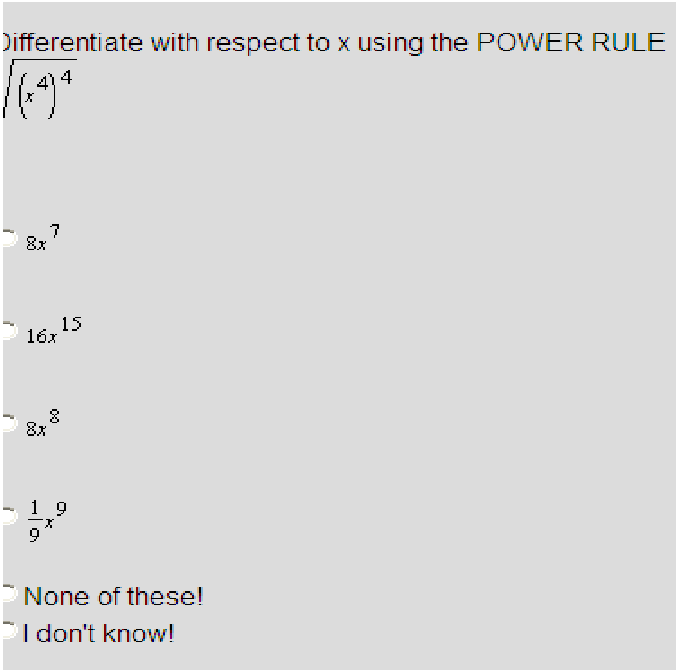


Figure 7.13: “sqrt((x^a)^b) MC” screenshot of question.

Looking carefully through the coding and the associated answer files, it is clear that there are 4 major error groups (mal-rules) encoded to diagnose student weakness, and are listed below with relative frequencies of response:

Correct / Mal-	Error	Sqrt (x^a)	Sqrt((x^a)^b)	Total	Percentage
Correct	Nil	298	307	605	82.3
Ignore square	PS	17	14	31	4.2
Wrong power	US	20	9	29	3.9
Integrated	PS	5	3	8	1.1
None of the	PS	30	23	58	7.9
Other	PS	2	2	4	0.5

Table 7.14: Mal-rules for differentiation of algebraic functions questions EC1005.

Clearly on such lower cognitive level questions, the teaching aim for those who make mistakes is to remember the basic rules (as hinted in the question by the explicit cue “POWER RULE”); firstly to turn the expression into an index expression, and then to remember how to differentiate the power of x correctly. Thus it is a multi (2)-structural question, and both processes could yield uni-structural mistakes (as in the wrong power (+1) mistake, or the other pre-structural mistakes (integrated, ignore square-root, none of the above). The relatively high frequency (7.9%) of “None of the above”, clearly suggests that students were finding other answers to those listed, or just guessing. My work in table 4.6 shows just how many other evidenced-based mal-rules could have been used for the construction of this question, and will remain a development point.

On middle skill questions, we see the facility values range much more (0.659 to 0.871), and looking at a selection of MC questions from use of the product rule on exponential functions, we see that the questions involve both the use of product and (to a lesser extent) chain rules. As described previously with the skill group: Differentiation / Chain Rule / Binomials, students have generally struggled more with this skill than others.

	Cubic*Exp(cubic)	Quadratic*Exp(ax)	Quadratic*Exp(cubic)	Quadratic*Exp(linear)	Quartic*Exp(quadr)
Facility	0.727	0.864	0.659	0.871	0.773
Discrimination	0.542	0.521	0.557	0.583	0.608

Table 7.15: Facility and Discrimination indices for middle cognitive questions

The following example in figure 7.16 is typical of such a question, which requires two rules (and thus is a multi-structured problem) to be correctly applied to yield the correct response.

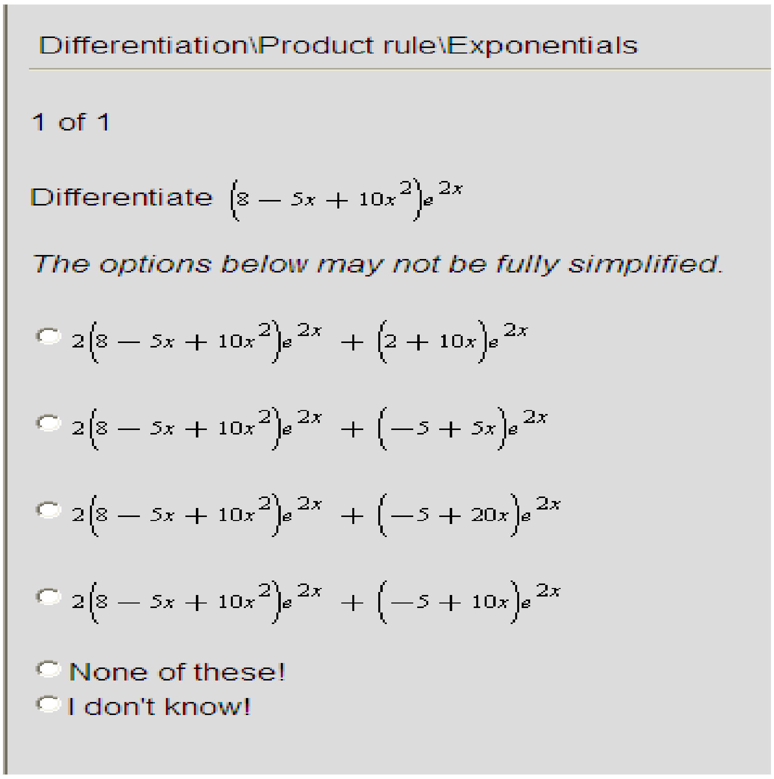


Figure 7.16: Differentiation of exponentials using product and chain rule for EC1005.

The structure of the mal-rules and distracters in the answer boxes is that of a full product rule solution (i.e. the question has a product of two terms and all solutions have two products of two terms), with generic chain rule form (i.e. the coefficient in front of one term). These explicit cues for the students should show them that both product and chain rules need to be used, so errors are likely to be consigned to uni-structural form, rather than pre-structural form (“None of these”).

In this question, all the candidates need to do in reality is identify the derivative of the quadratic function and select that option based on the secondary answer term, as the primary term is identical for all options. The mal-rules are based on uni-structural mistakes candidates have shown themselves likely to make on differentiation of polynomial functions here, rather that testing the nature of the product or chain rules. A weak candidate may indeed spot that (-5 + 20x) is indeed the correct derivative and thus go for this answer without recourse to the product or chain rule. If this is the testing aim of the question, then this question would be better off not in MC format, or be written with a few more distinct distracters to test the uni-structural mistakes from product rule (or chain rule) usage. As such, the nature of ease in obtaining the correct solution goes most of the way to explaining why it has a low discrimination index (0.521).

The very nature of writing suitable MC responses is key to completing the testing aims of these questions. Arriving at distracters that are too similar to each other or to the answer is liable to lead the students to choose the correct answer disproportionately often or infrequently, so a fuller evaluation of the choice of distracters in these similar cognitive level questions will allow not only to evaluate likelihood of specific student errors, but to test the efficacy of diagnosis of errors (i.e. whether the spread of distracters is wide enough and helpful enough in the questions to allow confidence in the spread of frequencies of each choice).

	Cubic*Exp(cubic)	Quadratic*Exp(ax)	Quadratic*Exp(cubic)	Quadratic*Exp(linear)	Quartic*Exp(quadr)	Totals	%
Correct	101	133	85	108	102	529	77.6
D1	3	7	7	4	4	25	3.7
D2	18	2	23	4	12	59	8.7
D3	6	5	5	3	4	23	3.4
D4	4	0	3	0	3	10	1.5
None of these	7	6	6	3	6	28	4.1
Do not know / Blank	2	2	0	2	2	8	1.2

Table 7.17: Results by question for product rule / exponential questions for EC1005.

On the answer files, each distracter (leading to a specific mal-rule) is labelled merely D1 to D4, and the answer "None of these" will feature as the correct answer randomly 10% of times. This will fulfil **Gill & Greenhow (2006)'**s twin aims of making sure that students don't guess all of the time, and also that they develop confidence in their answers rather than relying on them appearing all of the time in the options. The "None of these" figures above are for those candidates who incorrectly chose this answer, when the correct one was sitting above that. For each distracter, there is a 2.5% chance that it will not appear (as with “None of these” being correct), and in most cases, the incidence of each distracter being chosen is quite small. However, the significant observation is that the selection of D2 is highest of all in Questions 1, 3 and 5 (which are the ones with lowest facility values).

Looking in more depth at the first question: “Differentiate cubic * Exp (cubic)”, we can see how the distracters were formed and how the question is testing which skills:

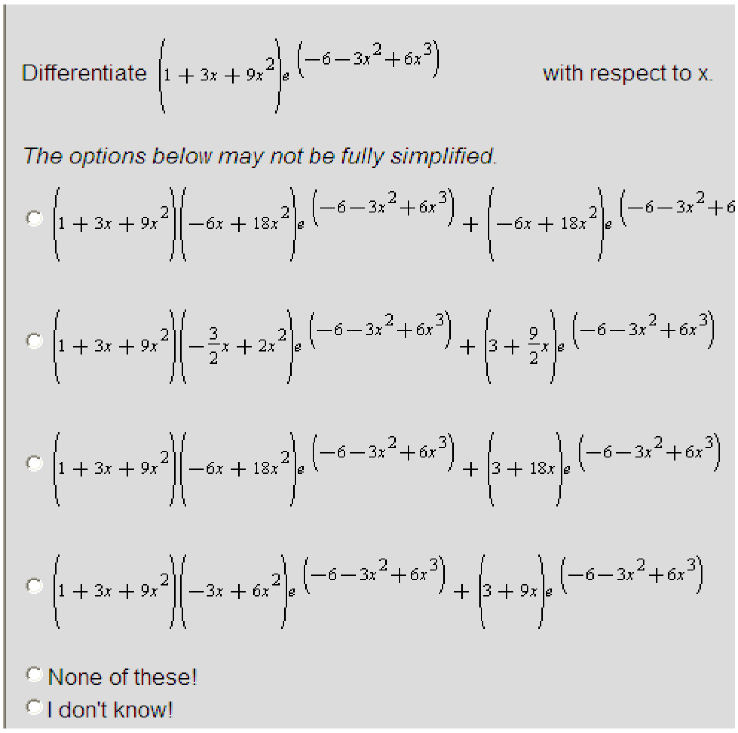


Figure 7.18: Differentiate cubic * Exp (cubic) screenshot for EC1005.

Here, again the structure of all four solutions is identical, i.e. the testing process isn’t looking to test whether candidates can use the formulaic process of the product rule, or the format of the chain rule, but merely can they differentiate two cubic expressions correctly. The associated errors built into the distracters are merely those found in differentiating basic polynomials. For each distracter, the coding is based on supposed mistakes that candidates are likely to make, **Greenhow et al. (2006)**:

Answer	1	2	3	4
Distracter	D2	D3	C1	D1
Description	Correct first half, but used derivative of cubic in 2nd half instead of derivative of quadratic.	Used division instead of multiplication for derivative of ax^b: Uses a/b x ^ b-1 instead of ab x ^b-1	Correct	Differentiates the polynomials wrongly, i.e. no multiplication. Chooses: ax^b becomes ax^b-1 instead of ab x ^b-1

Table 7.19: Distracter definitions for “Differentiate cubic * Exp (cubic)” for EC1005

Distracter 4 (only used when the answer is “none of these”), has the same correct format as C1, but includes a rogue “+C” inside the brackets of the differentiated terms. The lack of incidence partly explains why it is hardly ever chosen, and equally the “+C” should trigger alarm bells for candidates on a differentiation question as it stands out differently from the other questions. Distracter 2 has the same generic description across all three harder middle skill questions as in table 7.19 above, and is chosen more often than all other wrong answers in these questions. Candidates are scanning the solutions looking for the right structure. Once they spot that all answers have the same structure they may well be looking for commonalities, i.e. they see both D2 and C1 answers with the same first half. Mentally they are now likely to discount the other two answers. In both of these, they can scan across and see that there is commonality of the terms in the 2nd half of the D2 answer, which must be appealing from work done with chain rule questions in the past. Thus the distracter has itself become very distracting, without really identifying a particular mal-rule.

While the other distracters are well chosen, and would have tripped up those who have yet to grasp some polynomial differentiation, the nature of these mal-rules is such that students are likely to take the easier route and scan solutions before attempting the question. Thus our testing aim of understanding and application of the product or chain rules is not being met.

When looking at product rule questions based purely on polynomials (thus no chain rule to complicate matters), we see a strikingly different set of results (despite being all MC question types):

	Cubic*Cubic	Quadratic*Cubic CubibExp(ax)	Quadratic*Quarti	Quartic*Quartic	Cubic*Quintic	Total	%
Facility	0.606	0.681	0.723	0.677	0.813		
Discrimination	0.672	0.648	0.637	0.584	0.577		
Correct	86	96	112	86	109	489	69.9
D1	7	10	5	12	4	38	5.4
D2	11	11	9	11	8	50	7.1
D3	8	6	12	8	1	35	5.0
D4	0	0	1	0	2	3	0.4
None of these	28	16	15	10	10	79	11.3
Do not know	2	2	1	0	1	6	0.9

Table 7.20: Results by question for product rule / polynomial questions for EC1005.

Not only are these questions better differentiators with the discrimination index (0.624 compared to 0.562), they are also quite a bit harder given the lower facility levels (69.9% correct compared to 77.6% correct). The distracters don't seem to attract attention anything like as much as the exponential product rule questions and mistakes are liberally spread between D1, D2 and D3. What really jumps out is the level of mistakes by wrongly selecting "None of these" (11.3%).

Looking further at "Cubic * Quadratic" questions, the type of distracters used are typical of all of these types of questions.

Caroline is trying to differentiate

$$\left(-8+2x-10x^2+4x^3\right)\left(2-x-6x^2\right)$$

with respect to x. After applying the product rule and multiplying out, she should get:

- $\left(-16+164x-156x^2+232x^3-200x^4+48x^5\right)+\left(-2-23x+18x^2+72x^3\right)$
- $\left(8+94x-14x^2+116x^3-48x^4\right)+\left(-20x+26x^2+52x^3-48x^4\right)$
- $\left(8+94x-14x^2+116x^3-48x^4\right)+\left(4-42x+32x^2+108x^3-72x^4\right)$
- $\left(-16+84x-40x^2+108x^3-40x^4\right)+\left(4-42x+32x^2+108x^3-72x^4\right)$
- None of these!
- I don't know!

Figure 7.21: Screenshot of cubic * quadratic question for EC1005.

We can see that the first answer jumps out at us as different from the rest given the powers of the last term in each bracket, and thus intuitively warrants little further attention. From pattern recognition, we can equally see that the first term is repeated in answers 2 and 3, and the last term is repeated in answers 3 and 4. Thus, without any working, it would be instinctive to select answer 3 without completing any formal checking or indeed solving in the first place. Unfortunately, this strategy would yield the correct answer in this instance.

At least if we had the sense to check simply by multiplying the last terms of each bracket yields $-24x^5$, and when differentiated this gives $-120x^4$. A further quick check along each of the answers shows that answer 3 is also the only one to sum to $-120x^4$. The dangers of answer spotting are very clear in these questions. The same logic doesn't work every time, as there is a scenario encoded here whereby all four answers are wrong, and there are two matching pairs of first term and last term, overlapping with one (wrong) answer.

Answer	1	2	3	4
Distracter / Correct	D2	D3	C1	D1
Description	Wrong product rule: Uses v dv/dx + u du/dx instead of: v du/dx + u dv/dx	Correct first term (from u dv/dx), but the 2nd term is wrong. The product v du/dx has the wrong du/dx: ($-10x + 8x^2$) instead of ($2-20x+12x^2$), multiplying by the new power not the old one, and forgetting the 2 from 2x.	Correct	Correct second term (from v du/dx), but the 1st term is wrong. The product u dv/dx has the wrong dv/dx: (2 -10x) instead of (-1-12x), mis-multiplying the quadratic term, and forgetting the 2 disappears

Table 7.22: Distracter definitions for Differentiate cubic * quadratic question for EC1005

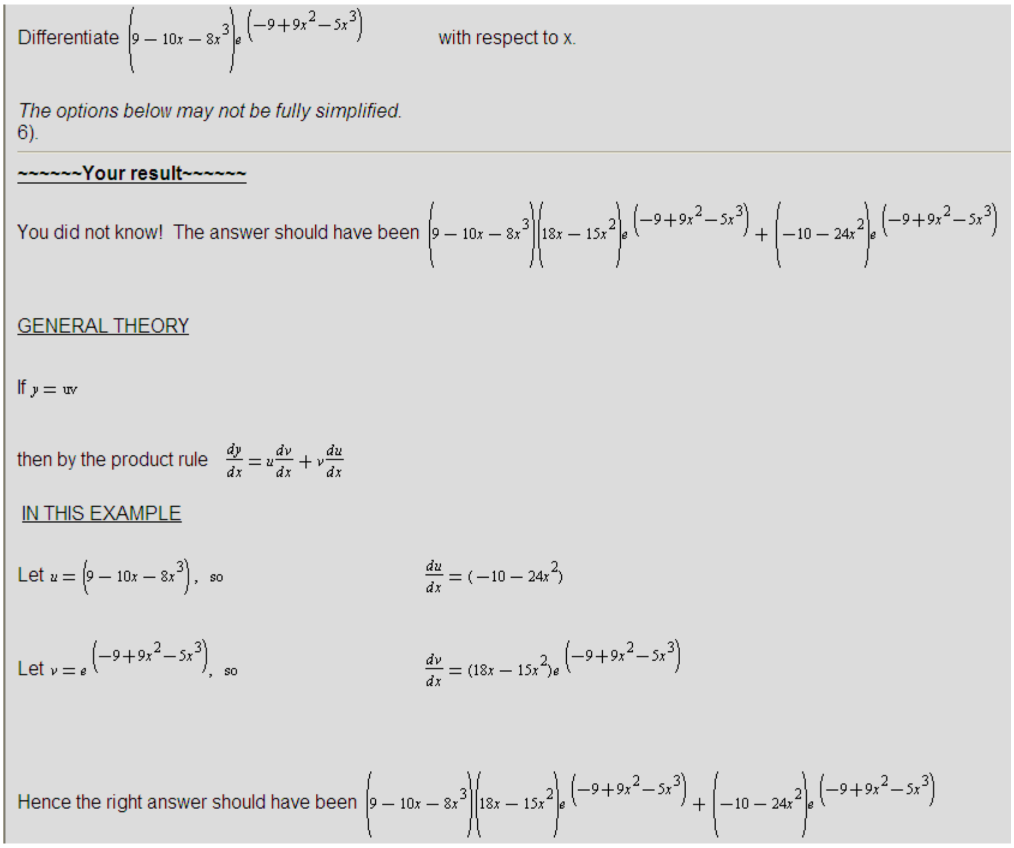
Looking critically at this question and others in the product rule / polynomial group, I can see how many students could “guess” the correct answer, or indeed use the sensible check of the final term to isolate the only possible correct answer. How many would then check the full working to make sure it was correct is hard to work out, but I suspect rather few. The distracters here (including D4 which is a variation of D1) don’t inspire likely choosing given the nature of the mistakes seeming unusual (as evidenced from table 4.6).

A more sensible plan to me would be to use D2 (wrong way round for the product rule) as our primary uni-structural mal-rule, and then to use other distracters based on differentiating the quadratic / cubic without one term (again obtained via a uni-structural mistake), or using the logical evidenced based mal-rules from table 4.6 whereby polynomial differentiation mistakes are incorporated as uni-structural mistakes within the product rule format, i.e. where: ax^{b} goes to ax^{b-1} instead of $ab\,x^{b-1}$ or ax^{b} goes to $a/b\,x^{b-1}$ instead of $ab\,x^{b-1}$.

The fiddly nature of these questions also makes it very hard to correctly arrive at the exact answer, and even one term out would lead a student to select “None of these” almost certainly wrongly, which explains the high incidence of these choices (11.3%). They are a better discriminator than the other product rule questions, but there is huge mileage in re-writing each set of questions with evidence-based mal-rules (or as RNI questions) to cope with more likely errors, and also to select mal-rules that don't yield an intuitively guessable solution from the selection of options.

7.6 Feedback

All that remains to be added in this results section is a thought about the feedback for students on these questions. Here below, showing a screenshot of the feedback page (following a wrong answer), we see the standard pattern that the question is given, followed by the correct answer. Then follows some general theory (including the product rule formula), and the intermediate calculation steps, showing the derivatives of each function – to high-light possible errors), before putting the steps together at the end.



Differentiate $\left(9-10x-8x^3\right)e^{\left(-9+9x^2-5x^3\right)}$ with respect to x.

The options below may not be fully simplified.
6).

~~~~~~Your result~~~~~~

You did not know! The answer should have been $\left(9-10x-8x^3\right)\left(18x-15x^2\right)e^{\left(-9+9x^2-5x^3\right)}+\left(-10-24x^2\right)e^{\left(-9+9x^2-5x^3\right)}$

GENERAL THEORY

If $y=uv$

then by the product rule $\frac{dy}{dx}=u\frac{dv}{dx}+v\frac{du}{dx}$

IN THIS EXAMPLE

Let $u=\left(9-10x-8x^3\right)$, so $\frac{du}{dx}=\left(-10-24x^2\right)$

Let $v=e^{\left(-9+9x^2-5x^3\right)}$, so $\frac{dv}{dx}=\left(18x-15x^2\right)e^{\left(-9+9x^2-5x^3\right)}$

Hence the right answer should have been $\left(9-10x-8x^3\right)\left(18x-15x^2\right)e^{\left(-9+9x^2-5x^3\right)}+\left(-10-24x^2\right)e^{\left(-9+9x^2-5x^3\right)}$

Figure 7.23: Screenshot of feedback after a product rule / exp question for EC1005.

Such a piece of feedback can only serve as very helpful for improving students scores, but it depends entirely on:

- Whether the students look at feedback.
- Whether they understand the basic theory enough to follow the solution (printing off for future retakes).
- Whether they can then translate the theory into analysing their own working to see where they went wrong.
- Whether they can then follow the theory and worked example with a new question that has the same structure but different numbers enough times to be confident of doing the question again without notes.
- Anecdotal evidence suggests that students say they learn maths from this feedback rather than from books or lecture notes.

7.7 Conclusions

- Students may be good at guessing / eliminating some MC answers if they give away too much of the correct answer
- Students won’t use “Don’t know”, but will use “None of these” often for fiddly questions

More work is needed to iron out likely differentiation mistakes, particularly at the basic skill end (polynomials), to make questions better differentiators at this level and also to help write questions at the middle and higher cognitive levels.

Chapter 8. Analysis of CAA tests on Integration

8.1 Integration

As shown in chapter 7, for Economics students, a sound understanding of the foundations of Calculus is key to understanding future economic courses. In the previous chapter, I focussed on differentiation questions in part to help answer whether students made the same mistakes on paper-based tests as online, and to look temporally to see whether questions became harder or better discriminators over time with increased teaching resources and preparation. I showed that more work is needed to iron out likely differentiation mistakes, particularly at the basic skill end (polynomials), to make questions better differentiators at this level and also to help write questions at the middle and higher cognitive levels. Our major question, “do students make the same mistakes on paper as online?”, was also informative as we showed that students could easily make different mistakes online when faced with a set of distracters that looked unhelpful to them.

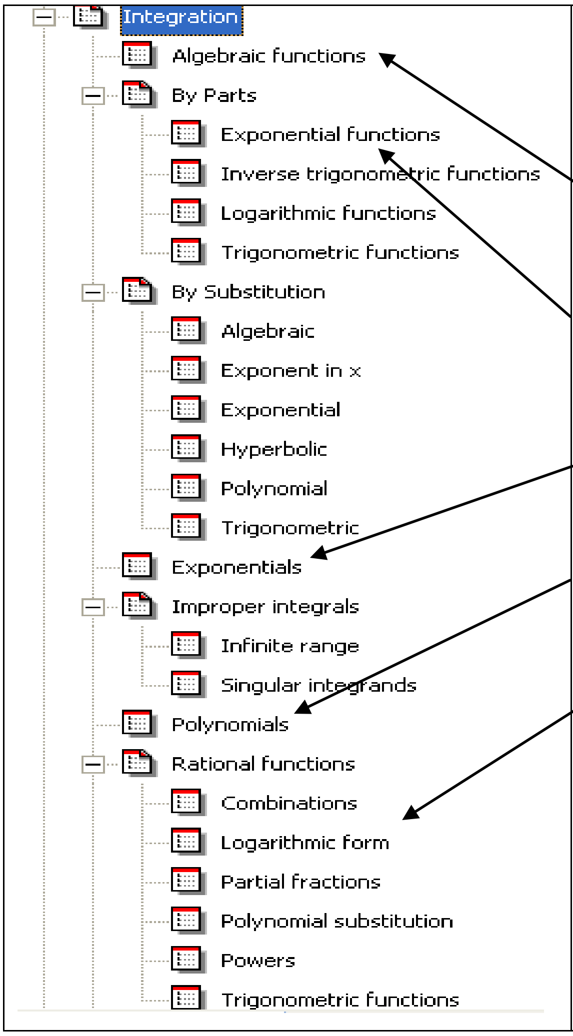
Equally students are inclined to guess the answer if the distracters give away too many explicit cues, or are too similar to each other to suggest elimination. Students too may well select “None of these” too often when the distracters were not well based on evidence of previous mistakes (a posteriori) and their attempts which resulted in a common uni-structural mistake didn’t appear in the likely list of distracters. Anecdotal evidence from Nottingham University suggests that students who fail to check answers will often choose “None of these” when they don’t see their first solution present. I am still keen to see CAA develop tests spanning a range of skills, not just for diagnosis purposes, but also to develop accurate and reusable formative assessments. Such careful analysis of efficacy of questions and mal-rules is thus vital for the future use of these tests.

A sample of integration questions was tested on Economics first year students (course code: EC1005) at Brunel University during 2008-2009, **Greenhow (2010)**. The Economics students sitting level 1 (a compulsory course in the 1st semester) are largely expected to achieve 300 UCAS points for entry to the course, and level 1 students must have at least AS mathematics course (thus incorporating many foundation skills in calculus). About 80% of 2008-2009 students have achieved A Level mathematics, in roughly equal proportions of each grade A to E, **Greenhow (2009)**.

The objectives of these tests were as follows:

- To diagnose basic calculus deficiencies at an early stage, and allow for correction
- To homogenise the mathematical levels of the cohort (who had a range of backgrounds in maths from GCSE through to A-Level)
- To understand basic integration rules and apply them to integrate polynomials and algebraic functions (basic level skills)
- To understand intermediate integration rules and apply them to integrate exponential and logarithmic functions (intermediate level skills)
- To understand integration by parts and apply it to harder questions on function products including exponential functions (advanced level skills)

769 tests (see Appendix 6) were generated from the EC1005 "Integration ASSESSMENT test", of the 2008-2009 cohort, **Greenhow (2009)**. The 32 questions for the test were selected from the Mathletics suite of questions, as detailed in figure 8.1:



The following topics had been included in the EC1005 course:

Integration\Algebraic Functions

Integration\By parts\Exponential Functions

Integration\Exponentials

Integration\Polynomials

Integration\Rational Functions\Logarithmic form

Table 8.1: Integration topic areas in Question Manager

8.2 Learning objectives

Within the set of integration questions, I plan to use of an assessment taxonomy to classify questions by focussing on the cognitive skills involved in each. Expanding **Bloom (1956)'s** taxonomy, namely: Remember, Understand, Apply or Analyse, can categorise most questions in mathematics well, **Baruah (2007).** However it can sometimes fail to describe accurately certain questions or tested skills. Integration is one topic whereby each skill in this test is built on foundations of basic integration and polynomial manipulation, and we can easily appreciate how these questions get harder and more multi-structural by building on understanding and recollection of ideas from a basic level.

I have also designated a cognitive level for questions that required a much more advanced "Problem Solving" ability often required and tested in tertiary level mathematics, as a mixture of Bloom's "Analyse" , "Synthesise" and "Evaluate" skills at the top of the learning objectives pyramid.

For these questions, one standard approach will not suffice, as there will be some initial simplification or expansion needed in the question so that it is re-written as either two questions, each of which can be solved in a standard way, or the problem is reduced to a simpler problem by means of a "trick", that requires some conscious creativity (a "relational" problem as defined under the SOLO taxonomy).

The learning objectives of these integration tests were categorised according to four cognitive levels as follows:

1) **Basic level** (Remember):
a) Integration being "the opposite" of differentiation,
b) Indefinite integrals of polynomial and exponential functions.

2) **Intermediate level** (Remember and Understand):
a) Indefinite integrals of rational functions (using established rules),
b) Definite integrals of polynomial and exponential functions, evaluating and simplifying answers.

3) **Advanced level** (Remember, Understand and Apply)
a) Integration by parts general formula, choosing which function to differentiate,
b) Simplification ideas, or a known general formula for integrating (f'(x) / f(x)) resulting in a logarithm form or the application of symmetry to an integral with limits from –S to S.

4) **Super advanced level** (Remember, Understand, Apply and Problem solve)

a) To apply prior learning skills to new integration problems, including simplification of a rational function before integrating (e.g. use of partial fractions),
b) Changing a (new) product of two functions into separate functions each of which could be integrated separately.

In these questions (in the five main integration topics), there are no questions which require modelling of a situation or analysis of a physical situation for forming and then solving such problems.

Of the 32 different questions used in this analysis (detailed below in table 8.1), I have designated them into the five different skills, and each skill has been allocated one or more of the four cognitive levels I proposed in section 8.2 using in this analysis:

Skill	**Cognitive level required**
Integration\Algebraic functions	Basic
Integration\By Parts\Exponential functions	Advanced / Super Advanced
Integration\Exponentials	Intermediate
Integration\Polynomials	Basic / Intermediate / Advanced
Integration\Rational functions\Logarithmic form	Advanced

Table 8.2: Classification of skills and difficulty of integration questions

In classifying the questions by cognitive levels, I used the definitions at the beginning of this section, i.e. **Integration\Algebraic functions** are all basic cognitive levels, as they just require remembering of rules and are all indefinite integrals.

Similarly, **Integration\Exponentials** are intermediate cognitive level questions, as they require memory and understanding of integration rules and are all definite integrals.

Integration\Rational functions\Logarithmic are all questions at the Advanced cognitive level, as they require memory, understanding and application of integration rules.

Primarily, I designated the skill of **Integration\Polynomials** is at a basic cognitive level for indefinite integrals, and intermediate cognitive levels for those questions which require definite integrals. The one exception to this was the indefinite integral question from –S to S, which could be achieved using symmetry of the integrand, and thus tagged it at an Advanced cognitive level.

Integration\By Parts\Exponentials is tagged at the Advanced cognitive level also by definition, but the structure of three questions, I designate Super Advanced cognitive level:

(Ax+C)exp(bx); MC	Integration\By Parts\Exponential functions	Super
(x^2-Ax)e^x; NI	Integration\By Parts\Exponential functions	Super
Int(exp(Bx)cos(Cx),x), B,C +ve; MC	Integration\By Parts\Exponential functions	Super

e.g.: $\int (ax+c)e^{bx}dx$ requires the splitting of the product, or careful integration by parts.

e.g.: $\int (x^2 - Ax)e^x dx$ requires the application of integration by parts twice.

e.g.: $\int e^{bx}\cos(Cx)dx$ requires the application of integration by parts twice.

Question Description	Topic	Cognitive level	Times presented	Times answered	Mean (facility)	Difficulty	Standard Deviation
int b x^(1/q); q +ve; MC	Integration\Algebraic functions	Basic	232	231	0.887	0.113	0.317
int b x^(1/q); q -ve; MC	Integration\Algebraic functions	Basic	264	262	0.874	0.126	0.332
int b x^(p/q); q +ve; MC	Integration\Algebraic functions	Basic	243	243	0.877	0.123	0.33
(Ax+C)exp(bx); MC	Integration\By Parts\Exponential functions	Super	59	58	0.621	0.379	0.489
(x^2-Ax)e^x; NI	Integration\By Parts\Exponential functions	Super	72	72	0.694	0.306	0.959
Axexp(Bx); MC	Integration\By Parts\Exponential functions	Advan	63	62	0.758	0.242	0.432
int((x/a)(exp(x/q+p)),x,a,q,p); MC	Integration\By Parts\Exponential functions	Advan	75	75	0.507	0.493	0.503
int(ax(exp(px)),x,a,p); MC	Integration\By Parts\Exponential functions	Advan	74	74	0.514	0.486	0.503
int(ax(exp(px/q)),x,a,p,q); MC	Integration\By Parts\Exponential functions	Advan	73	72	0.528	0.472	0.503
int(bx(exp(x/q)),x,b,q); MC	Integration\By Parts\Exponential functions	Advan	77	77	0.455	0.545	0.501
Int(exp(Bx)cos(Cx),x), B,C +ve; MC	Integration\By Parts\Exponential functions	Super	6	6	0.167	0.833	0.408
x*exp(bx); MC	Integration\By Parts\Exponential functions	Advan	81	81	0.679	0.321	0.47
x^2exp(px); MC	Integration\By Parts\Exponential functions	Advan	72	71	0.38	0.62	0.489
int(exp(px),x,a,b); MC	Integration\Exponentials	Inter	185	185	0.703	0.297	0.458
int(exp(px/q),x,a,b); MC	Integration\Exponentials	Inter	176	176	0.597	0.403	0.492
int(exp(x/q),x,a,b); MC	Integration\Exponentials	Inter	182	182	0.582	0.418	0.495
Int(x^n+exp(-mx),x,A,B), A, B +ve; NI	Integration\Exponentials	Inter	172	172	0.459	0.541	0.5
int(a+bx+cx^2,x); MC	Integration\Polynomials	Basic	89	88	0.886	0.114	0.319
int(ax^b,x); MC	Integration\Polynomials	Basic	66	66	0.848	0.152	0.361
int(ax^b,x,0,1); MC	Integration\Polynomials	Inter	65	65	0.846	0.154	0.364
int(ax^b,x,l,u)... 0 lt l lt u; MC	Integration\Polynomials	Inter	78	78	0.859	0.141	0.35
int(ax^b,x,l,u)... fractional l & u; MC	Integration\Polynomials	Inter	88	88	0.886	0.114	0.319
int(ax^b,x,l,u)... l lt 0 lt u; MC	Integration\Polynomials	Inter	62	62	0.71	0.29	0.458
int(ax^b,x,l,u)... l lt u lt 0; MC	Integration\Polynomials	Inter	70	70	0.771	0.229	0.423
int(ax^b,x,l,u)... u lt 0 lt l; MC	Integration\Polynomials	Inter	87	87	0.816	0.184	0.39
int(ax^b,x,-s,s); MC	Integration\Polynomials	Advan	76	76	0.842	0.158	0.367
int(ax^b+cx^d,x); MC	Integration\Polynomials	Basic	78	78	0.897	0.103	0.305
int(ax^(n-1)/(bx^n+g),x); MC	Integration\Rational functions\Logarithmic	Advan	135	135	0.681	0.319	0.468
int(Ax^(n-1)/(Bx^n+G),x,a,b), a>b +ve	Integration\Rational functions\Logarithmic	Advan	118	115	0.539	0.461	0.501
int(Ax^(n-1)/(Bx^n+G),x,a,b), a > b -v	Integration\Rational functions\Logarithmic	Advan	138	133	0.436	0.564	0.498
int(Ax^(n-1)/(Bx^n+G),x,a,b), a,b +ve;	Integration\Rational functions\Logarithmic	Advan	126	126	0.5	0.5	0.502
int(Ax^(n-1)/(Bx^n+G),x,a,b), a,b -ve;	Integration\Rational functions\Logarithmic	Advan	150	145	0.476	0.524	0.501

Table 8.3: Integration questions for EC1005 students categorised by summary statistics.

8.3 Outcomes from the tests

The 32 questions are listed in full with outcome proportions in Appendix 7. Using the Integration tab on the EC1005 (**EC1005_CAA_analysis_stats.xls**), we can see that each question (aside from one anomaly that occurred 6 times) presented itself between 59 and 243 times (mean of 110.8), with 3532 questions presented in all. Given there were 769 tests realised, then the students sat an average of 4.593 questions each.

Of the 32 types of questions, 26 of them are Multiple Choice (MC), 4 are of the form: Responsive Numerical Input (RNI) and 2 requiring a Numerical Input (NI). As described in chapter 7, it is very difficult to draw meaningful comparisons (aside from facility level) for outputs for NI question formats as the responses (incorrect) are not recorded, but it can still be useful for comparisons between the discrimination values for the question types.

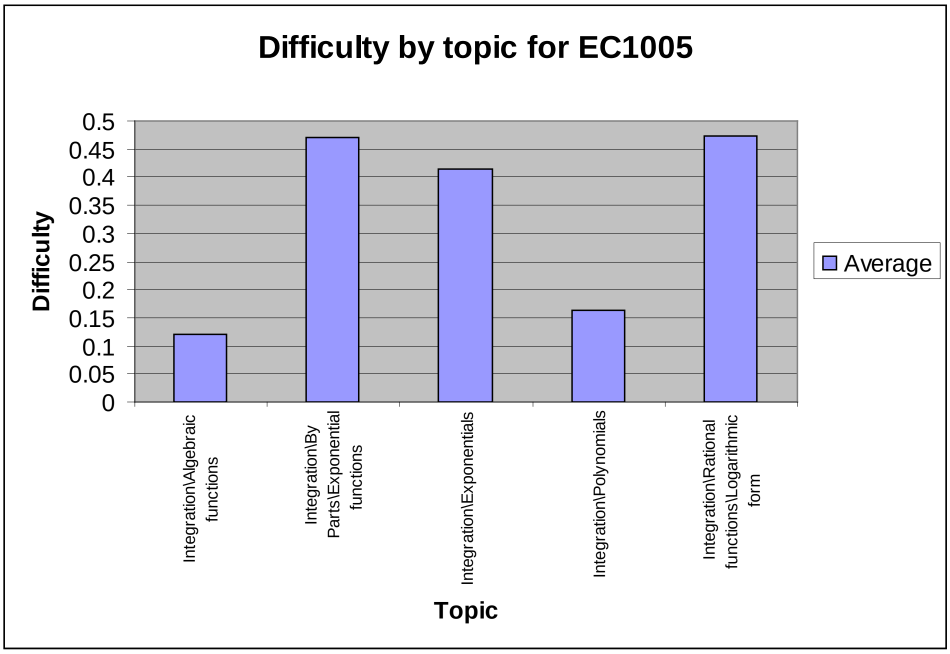
It is worth pointing out again that the facility value is calculated from the total number of correct responses out of the total number of submitted answers (rather than the number of times the question was presented, as some candidates may default on a question, or have IT issues so not complete that question). Overall, given that 3511 responses were given out of 3532 presentations (99.4%), there is scope for further research whether student default or IT issues or other reason accounts for this figure.

Comparisons of performance over the different skill areas leads to very interesting results for the EC1005 cohort. Looking purely at the average difficulty values for the topic areas, we find three skills which are well set and are a medium challenge to students (0.5 < facility < 0.6), and two skills which have a very low difficulty level and hence high facility value : (0.8 < facility < 0.9) as detailed in table 8.4 below.

Skill	**Facility**
Integration\Algebraic functions	0.8793
Integration\By Parts\Exponential functions	0.5303
Integration\Exponentials	0.5852
Integration\Polynomials	0.8361
Integration\Rational functions\Logarithmic form	0.5264

Table 8.4a: Summary statistics for facility by skill for EC1005 students on Integration

As a comparison, we define **Difficulty as 1 – Facility value**, and the summary statistics for difficulty shown in figure 8.3b very clearly indicate the relative difficulty of these skills for students:

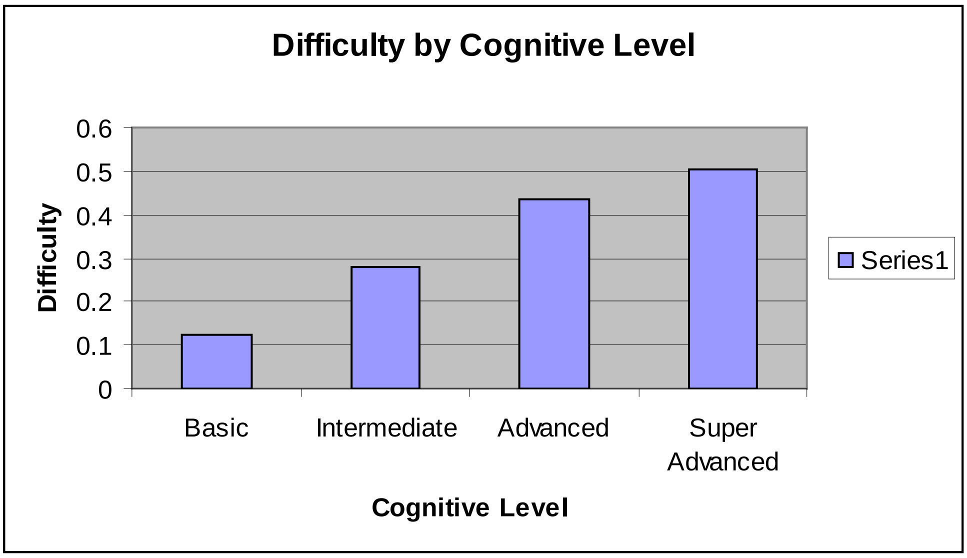


Graph 8.4b: Graphical summary for difficulty by skill for EC1005 integration

8.4 Comparisons between the cognitive levels

Cognitive Level	Basic	Intermediate	Advanced	Super Advanced
Difficulty	0.121833	0.2771	0.43676	0.506
Number of Questions	6	10	13	3

Table 8.5a: Cognitive Level difficulty statistics



Graph 8.5b: Cognitive Level difficulty statistics

If we suppose that there is a significant distinction between performance on questions requiring higher cognitive skills than lower ones, we can test for this difference by grouping the Basic and Intermediate questions into one cognitive cluster (1), and the Advanced and Super Advanced questions in another cognitive cluster (2).

Thus a null hypothesis, H_0 would state: There is no significant difference between performances on questions covering the lower cognitive skills to those covering higher cognitive skills. As opposed to an alternative hypothesis, H_1: There will be a significant difference between performance on questions covering the lower cognitive skills to those covering higher cognitive skills.

This is a non-directional (2-tailed) hypothesis test, and a reasonable assumption would be to look for a 5% significance level, so that we have a reasonable degree of confidence in our conclusions. We use a Mann-Whitney test, because is a non-parametric test for assessing whether two independent samples of observations come from the same distribution. In this case, we want to test whether or not the samples of lower cognitive questions are from the same distribution as the higher cognitive ones.

Following **Mann & Whitney (1947)**, we group observations (in this case the mean scores per question) into the two clusters and rank them against the whole set of 32 questions. Each cluster contains 16 questions, so we could consider this a paired test if we were confident the samples came from the same population (which we aren't). As the non-directional 5% level, we require the "U" (test statistic) to be 75 or more. The respective sum of ranks for group 1 and 2, R1, R2 respectively, with N1 and N2 (16 for the number in each group respectively and consequent U values for each cluster are given in table 8.6 below.

R1	R2
165	362
N1	N2
16	16
U1	U2
37	234

Table 8.6: Mann-Whitney U test values for Clusters 1 and 2 of EC1005.

Given that N is less than 20 for each sample, we won't have justification for using a normal approximation to the distribution, but the results are very clear nonetheless. Taking the lower U value (37), we see this is way below the critical value (as given in table 8.6 below), courtesy of: http://math.usask.ca/~laverty/S245/Tables/wmw.pdf

So we can reject H_0 in favour of H_1, that there is a statistical difference between the two samples, namely that there is a significant difference between performance on higher cognitive questions than lower cognitive ones. Even if the significance level of the test was reduced to 0.01 (non-directional), we still have a critical value for U of 60, given n1 = n2 = 16, so we would draw the same conclusion, only with a more stringent test and confidence interval. More work is needed for students preparing for test questions at a higher cognitive level.

Table A5.07: Critical Values for the Wilcoxon/Mann-

Nondirectional α=.05 (Directional α=.025)

	n_2															
n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42
11	-	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47
12	-	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53
13	-	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75

Table 8.7: Critical values for the Mann-Whitney U test at the 5% non-directional level.

8.5 Analysis of the quality of the basic level questions

Consider the skill area for which we find the difficulty to be very low (and requiring a basic cognitive level): "Integration\Algebraic Functions" below in table 8.7

Question description	Topic	Cognitive Level	Times presented	Times answered	Facility	Difficulty
int b x^(1/q); q +ve; MC	Integration\Algebraic functions	Basic	232	231	0.887	0.113
int b x^(1/q); q -ve; MC	Integration\Algebraic functions	Basic	264	262	0.874	0.126
int b x^(p/q); q +ve; MC	Integration\Algebraic functions	Basic	243	243	0.877	0.123

Table 8.8: Summary statistics for Integration\Algebraic Functions for EC1005

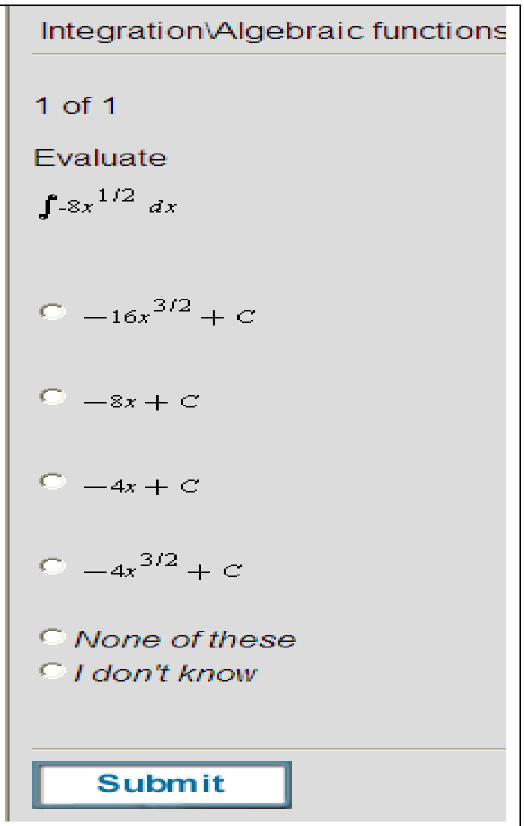
The question description: "int b x^(1/q); q +ve; MC" Refers to the following Multiple Choice question: Evaluate: $\int bx^{\frac{1}{q}}dx$ Where b and q are random integer parameters given	

Table 8.9: Screen shot for Integration of Algebraic functions

The nature of these questions is that if you can achieve one correctly, then you should (in theory) not have many problems with the others, as the processes for solution are generic across the question styles, and this fact is borne out by almost identical facility values in table 8.10, roughly 1% different, and standard deviations differ by less than 5% (0.317 to 0.332).

To compare further, I will consider the role of mal-rules and distracters in setting up and construction of these questions.

Question description	Topic	Times presented	Times answered	Mean score	Difficulty	Outcome analysis			
int b x^(1/q); q +ve; MC	Integration\Algebraic functions	232	231	0.887	0.113	**Outcome name**	**Times answered**	**Percentage answered**	**Mean for outcome**
						Correct	205	88.36%	68.69%
						not p+1	3	1.29%	19%
						diff & not p+1	1	0.43%	0%
						Diff	12	5.17%	32.25%
						Guess	1	0.43%	0%
						None Of These	8	3.45%	19.13%
						did not know	1	0.43%	0%
						Not Answered	1	0.43%	40%
int b x^(1/q); q -ve; MC	Integration\Algebraic functions	264	262	0.874	0.126	**Outcome name**	**Times answered**	**Percentage answered**	**Mean for outcome**
						Correct	229	86.74%	65.47%
						not p+1	10	3.79%	28%
						diff & not p+1	5	1.89%	16%
						diff	5	1.89%	4%
						guess	7	2.65%	28.57%
						None Of These	6	2.27%	22.83%
						Did Not Know	0		-
						Not Answered	2	0.76%	60%
int b x^(p/q); q +ve; MC	Integration\Algebraic functions	243	243	0.877	0.123	**Outcome name**	**Times answered**	**Percentage answered**	**Mean for outcome**
						Correct	213	87.65%	67.12%
						not p+1	4	1.65%	32.50%
						diff & not p+1	2	0.82%	10%
						diff	14	5.76%	34.50%
						guess	4	1.65%	0%
						None Of These	5	2.06%	32%
						Did Not Know	1	0.41%	0%
						Not Answered	0		-

Table 8.10: Outcome distribution for Integration of Algebraic Functions

The question: $\int bx^{\frac{p}{q}}dx$, where a) p = 1, q > 0, b) p = 1, q < 0 (not -1), and c) p anything, q > 0 refers to the three different classes of questions referred to in table 8.10.

The choices of correct answer are randomly rotated through the four distracters and "None of these" (which occurs 10% of the time), so that either 3 or 4 distracters can be used in the selection of outcomes. The feedback from the distracters and in the outcomes box above shows that the mal-rules are encoded as follows (outcome short-form from table 8.11 listed in quotes):

D1	You have not increased the power of x by 1 when doing the integration (“not p+1”)
D2	Two mistakes here. You appear to have differentiated the x term, not integrated it. You have not increased the power of x by 1 when doing the integration (“diff and not p+1”).
D3	You appear to have differentiated the x term, not integrated it. (“diff”)
D4	Your answer is wrong. Check your working against the solution to identify your mistake. (“guess”).

In analysing these mal-rules, it is important to look carefully at such CAA results above to see if any distracters never get selected, or indeed get over-selected, and to look carefully at questions where “None of these” is wrongly selected more commonly than any of the distracters. Above, in table 8.9, it is clear that the majority of (the admittedly small number of mistakes) are pre-structural in nature, and caused by selecting the “diff” option, some 31 out of 736 responses (4.2%). Here the student has merely differentiated the function (and selected an option similar to a derivative, despite the rogue “+C” floating after it).

The option “None of these” is selected wrongly 19 times out of 736 solutions (around 2.6%). This compares favourably with EC1005 differentiation responses in section 7.5, whereby 11.3% of wrong selections were for this option. Clearly this low figure for Integration questions on algebraic functions shows that the range of distracters is good enough to catch most mal-rules. However, the much higher facility value also indicates that distracters are seldom going to be selected anyhow, so we need to analyse much harder questions to see if this selection of “None of these” is similarly replicated.

However, in encoding such mal-rules to the MC answers, it is clear that the one answer without a “+C” at the end will spring alarm bells in students minds and rule itself out (and this type of question cannot thus test this recall skill), and also cause the student to remember the “+C” term. So if we were to offer integrands without a “+C”, and derivatives (without it), then this might allow us further scope for development of these questions by offering them a clear choice. Given likely mistakes with finding derivatives instead of integrating (as detailed in table 4.6), we could make the following additions to distracters to decrease question facility:

1) Correct derivative

2) The power term correct, constant term found by multiplying not dividing by p+1.

3) The dividing by (p+1) has been ignored.

4) The correct constant term in front but the power decreased by one.

8.6 Analysis of the quality of the intermediate level questions

There are 10 other questions on integrating Polynomials, 8 of them following the pattern: $\int_{l}^{u} ax^{b} dx$, where L and U conform to certain restrictions, signs, or fractions.

The other two questions are of the form: $\int ax^{b} + cx^{d} dx$, which builds on the same skill set and similar skill set to that required in table 4.6 for differentiation. However, it is the definite integral questions I wish to compare here, as they require advanced skills.

Question style	**Number of attempts**	**Facility**	**Difficulty**	**Most common wrong selection**
Indefinite	66	0.848	0.152	Distracter: No constant
Definite: L = 0, u = 1	65	0.846	0.154	None of these
Definite: 0 < L < U	78	0.859	0.141	None of these
Definite: L, U fraction	88	0.886	0.114	None of these
Definite: L < 0 < U	62	0.71	0.29	Distracter: Evaluation of limits
Definite: L < U < 0	70	0.771	0.229	Distracter: Evaluation of limits
Definite: U < 0 < L	87	0.816	0.184	None of these
Definite: L = -S, U = S	76	0.842	0.158	Distracter: Twice evaluation 0 – S

Table 8.12: Summary statistics for definite integral polynomial integration questions.

For the average facility between those questions (0.8495 without the two harder questions), the definite integral question where: L < 0 < U is nearly 20% harder that the rest, while the definite integral question where: L < U < 0 is around 10% harder that the overall facility mean without them. Thus there is a significant difficulty challenge for students being assessed on those two questions, such that parity must be questioned in tests that randomly select one or other of those questions to compare overall understanding.

Interestingly, the question which comes out easiest for the students is the definite integral with fractional limits:

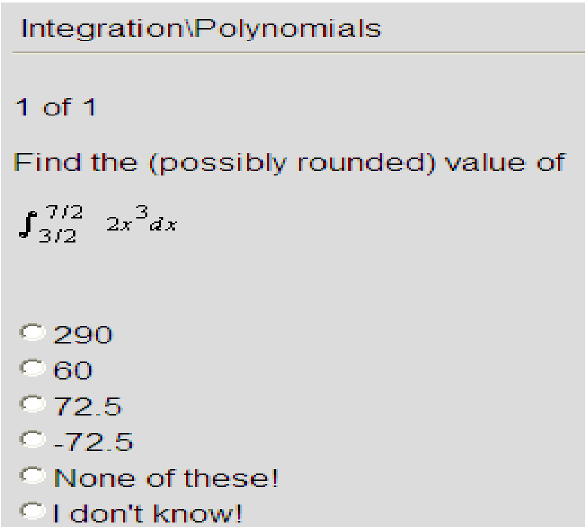


Figure 8.13: Screenshot of fractional limit definite integral of polynomial question.

The theory is that: $\int_l^u ax^b dx = \left[\frac{ax^{b+1}}{b+1}\right]_l^u = \frac{au^{b+1}}{b+1} - \frac{al^{b+1}}{b+1}$, so it should just be a small amount of "bookwork", and then evaluating using a calculator:

Here: $\int_{1.5}^{3.5} 2x^3 dx = \left[\frac{2x^4}{4}\right]_{1.5}^{3.5} = \frac{3.5^4}{2} - \frac{1.5^4}{2}$ = 72.5 (roughly)

So 72.5 is the correct answer from figure 8.13.

This multi-structural problem has many processes even before the use of a calculator to derive mal-rules based on uni-structural responses, so could yield many possible distracters. One reason this might be easier to evaluate than with integer limits is probably because a canny student could see that it is not going to produce a negative response (from the nature of the cubic curve for x > 0), and the fractional limits would likely yield a fractional answer, hence the one choice (72.5). The difference between this mean (0.886) and the next highest mean (0.859) is still around 3%, so there is reason to suggest that such distracters on this question to encourage some students to guess.

In this question the majority of student errors here were in selection of “None of these”, which tells us tellingly that none of the other options worked well as a distracter. This warrants further thought, as the level and choice of distracters is very clearly linked to the difficulty of the questions, so it is worth revisiting these mal-rules to make sure the questions are not overly easy or difficult owing to our choice of mal-rules and do encourage the students to work through the questions carefully without guessing.

The major distracters encoded into the definite integral questions were:

D1: Integrated but failed to divide by (b+1)
D2: Differentiated the expression (and evaluated with limits)
D3: Integrated but used b as a divisor, not (b+1)
D4: The negative of the actual answer.

In addition to these answers, I would add in a few other distracters:

1) Evaluation of the function with the limits (i.e. no integration or differentiation taken place)
2) Evaluation of integrand with “+” not “-“, i.e. F(u) + F (L), not F(u) – F(L)
3) Division of (b+1) replaced by multiplication of (b+1)
4) Evaluation of only the upper most limit, i.e. F(u)

It is only in the careful construction of these questions with well chosen distracters that we can achieve parity across the questions and evaluate errors made by students in order to tailor make their feedback. To leave 35 out of 103 (some 34%) of the wrong answers as either "None of these" shows up nothing of the students' mistakes, and renders the questions little more than NI format in those cases, indicative of guesswork. A more realistic MC question would, in my view, encode 5/6 distracters in with the main solution, which should allow very realistic diagnosis of the common student mistakes and misconceptions in the minority of cases (roughly 15% overall) who failed to understand this topic by making pre-structural mistakes or guessing.

8.7 Analysis of the quality of the advanced level questions

While basic polynomial integration seems to be grasped by a healthy majority of students sitting those questions, the skill of integration by parts (with exponential functions) is clearly one that causes many more errors and a much lower incidence of success. The topic: "Integration\By parts\Exponential functions" generates questions of the form: $\int (ax+c)e^{\left(\frac{px}{q}+d\right)}dx$, where some of c, d = 0 and some of a, p, q are 1.

Question style	Number of attempts	Facility	Most common wrong selection
q = 1, d = 0	58	0.621	Distracter: Integrated each term separately, formula incorrect sign.
c = 0, q = 1, d = 0	62	0.758	Distracter: Differentiating exp (bx) instead of integrating it
a = frac, c = 0, p = 1	75	0.507	Distracter: Chose exp(x) to be v instead of dv
c = 0, q = 1, d = 0	74	0.514	Distracter: Formula incorrect sign
c = 0, d = 0	72	0.528	Distracter: Formula incorrect sign
c = 0, p = 1, d = 0	77	0.455	Distracter: Formula incorrect sign
c = 0, a = 1, q = 1, d = 0	81	0.679	Distracter: Differentiating exp (bx) instead of integrating it and wrong sign in parts formula

Table 8.14: Question styles for Integration by Parts \ Exponentials

In every case bar one of the above formula, c = 0 and likewise d = 0, so the question should unravel like a standard integration by parts question, where the “By parts” strategy is used once if we carefully choose which function is our “u(x)”, and which will be our “dv(x)”. In fact, intuitively, it would appear that setting q = 1 would also greatly simplify the problem, but this is not proved in the two cases (question 2 and question 4 in table 8.14) above. Both realisations of the question look identical from the table, but the screenshots show the one key difference.

<table>
<tr>
<td>Using integration by parts evaluate
$\int 3x \exp(4x)\, dx$

$\exp(4x)(12x-3)+C$
$\frac{3\exp(4x)}{4}\left(x+\frac{1}{4}\right)+C$
$\frac{3\exp(4x)}{4}\left(x-\frac{1}{4}\right)+C$
$\frac{3x^2\exp(4x)}{8}+C$

None of these
I don't know</td>
<td>Evaluate
$\int 17xe^{4x}\,dx$

$-\frac{17}{4}e^{4x}\left[x+\frac{1}{4}\right]+C$
$\frac{17}{4}e^{4x}\left[x-\frac{1}{4}\right]+C$
$\frac{17}{4}e^{4x}\left[x+\frac{1}{4}\right]+C$
$17e^{4x}\left[x-\frac{1}{4}\right]+C$

None of these
I don't know</td>
</tr>
<tr>
<td>Facility = 0.758</td>
<td>Facility = 0.514</td>
</tr>
</table>

Table 8.15: Two almost identical realisations of “by parts questions”

In the question on the right, the skill is harder because there is no guidance or strategy given to the student (no explicit cues), whereas the one on the left, the student is clearly told to use “Integration by Parts”. Should this one explicit cue make such a significant difference? Or are the mal-rules different for the question as well? Looking more carefully in table 8.15, it does appear that any of the answers to the right hand question could be correct, so no distracters instantly remove themselves from likely selection. However, on the left hand side, one’s eye is drawn to the middle two answers to the question on the left as the 1st and 4th answer look like they are in a different form that "seems" wrong, or are too simple or missing something so could be discounted.

This student can go further in checking responses, in that by remembering the "by parts formula" has a middle negative sign, and the idea that we reduce order in products, specifically powers of x, so there is high likelihood of the 3rd answer (C) being correct on the left hand question without doing any calculation. In fact it turns out to be a very good hunch, as the correct answer would be (C) on the left question and respectively (B) on the right hand side, namely:

$$\int pxe^{qx}dx = \frac{pxe^{qx}}{q} - \frac{pe^{qx}}{q^2} + c = \frac{p}{q}\left(x - \frac{1}{q}\right)e^{qx} + c$$

Guesswork on the right-hand question would perhaps only knock out the one answer without a minus sign in it as distinct from the others. Could it be that these questions are hard enough to persuade students to guess more than we might expect them to? The analysis of outcomes from both questions yields 8.11% and 9.52% selections respectively for the option "None of these", which is more indicative of either pre-structural mistakes of unspecified form, or a poor guess.

Axexp(Bx); MC	Integration\By Parts\Exponential functions	63	62	0.758	0.758	**Outcome name**	**Times answered**	**Percentage of times answered**
						Correct	**47**	**74.60%**
						integrated each term separately	**2**	**3.17%**
						differentiated exp(Bx) instead of integrating	**4**	**6.35%**
						incorrect sign in the formula	**2**	**3.17%**
						differentiation instead of integration	**0**	
						None Of These	**6**	**9.52%**
						Did Not Know	**1**	**1.59%**
						Not Answered	**1**	**1.59%**
int(ax(exp(px)), x,a,p); MC	Integration\By Parts\Exponential functions	74	74	0.514	0.514	**Outcome name**	**Times answered**	**Percentage of times answered**
						Correct	38	51.35%
						formula incorrect, sign.	19	25.68%
						choose exp(x) to be v instead of dv	6	8.11%
						integral of dv wrong, sign	4	5.41%
						integral of dv wrong, constant	0	
						None Of These	6	8.11%
						Did Not Know	1	1.35%
						Not Answered	0	

Table 8.16: Outcomes for two integration by parts questions

8.8 Further discussion

The way students will learn how to eradicate errors on such questions would be to practise repeatedly onscreen until they were scoring above average success on similar type of questions. In order to draw comparisons and assess the categorisation of outcomes across cohorts, this chapter aimed firstly to categorise different topics and question styles for common skill requirements, and secondly to compare results within different topics in order to categorise difficulty tags for the questions for future cohort assessments.

Within those aims, I have unearthed yet more evidence that changing the mal-rules to a question, or adding in explicit introductory clues will change the facility of the question greatly. The other tangible conclusion here is that without properly structured mal-rules, the students are likely either to strategically eliminate options and head for a distracter or the correct answer a disproportionately high frequency, or else choose “None of these” rather more often than I would expect.

Feedback too will form the basis of further discussion as formulating effective feedback to the students is more and more key, and has a knock-on effect of being able to inform a larger trend / skill deficiency analysis for the cohort.

Chapter 9 – Conclusions and future direction

9.1: Online testing vs. paper-based testing

- Do students made the same mistakes on a paper-based test as an onscreen test?

If a question changes facility value greatly between a paper-based test and onscreen test, (and there are the same explicit cues given in the question stem), I have to ask myself: have I written the wrong mal-rules or written them in such a way as to overly distract a student from the correct answer by phrasing it in an unfamiliar way? Initial analysis, **Greenhow (2008)**, indicated that mal-rules play a large part in guiding students away from or towards specific answers.

Clearly the nature of selection of our mal-rules has changed the facility of Q5 on our algebraic diagnostic test considerably (section 5.2). Having changed from paper-based test to MC objective test, and 100 students sat the test, Q5 improved from being the hardest question (out of 16) to become only the 7th most difficult question, whereas Q15 had no change on the difficulty rank (12th out of 16). When looking further into this change on Q5, the pre-structural mistake of including an invisible bracket caught out many students (most common mistake) with a numeric paper-based response format but, when offered as a distracter on the MC test, it was clearly seen as wrong, and hardly selected at all. One preliminary conclusion suggests also that the nature of some multiple-choice answers will clearly cause the students not to follow that same mistake as if the question were NI format.

Looking further at the selection of paper-based mistakes in sections 5.2, and summarised from examiners' reports in table 4.6, it is clear that an equal number of pre-structural mistakes as uni-structural mistakes were made by students, and as such, we ought to include an equal balance of these as mal-rules for our distracters. However, virtually all mal-rules on CAA questions are encoded with uni-structural mistakes, which pre-suppose that students know some rudimentary theory on the topic tested, and also apply some process correctly even if on route to a wrong final response.

Greenhow & Gill (2004), and **Greenhow at al. (2002)** include a null ("none of the above" or "I don't know") choice as well. Such an option was not specifically available on paper-based tests, and so students who left an answer blank or arrived at a response not using a common mal-rule, would now have the option of selecting this option when sitting the onscreen test (without ever knowing how often that was the correct option). I focussed on looking carefully at the proportions selecting this option onscreen, and where the selection was above 5%, or it became the most popular wrong response, I concluded that the distracters were too poorly written to trap the likely mistakes made by students, and thus needed careful revision.

9.2: One diagnostic test fits all

- Can we devise a diagnostic test spanning skills?
- What are the limitations of the Excel worksheets as a diagnostic tool?
- Where will CAA improve on this?

Our MC objective C1 Excel diagnostic test can be sat with questions online and students submitting their responses via paper-based answer sheet. Its main limitation is that there is no facility to select responses onscreen to generate summary feedback onscreen after the end of the test. The process for designing such worksheets is also time-consuming, fiddly and the format is difficult to really make user-friendly.

Given that one major purpose of this thesis is to elicit useful error analysis and student feedback from diagnostic tests, then I can conclude that such formats are not appropriate for designing tests spanning large skill areas. However, CAA tests are reusable and the results data will allow much more refined topic analysis by errors and also allow tailor made tests for candidates / cohorts to be made to help improve on specifics.

Where this analysis is useful for teaching staff, is that they readily use non-CAA processes for generating feedback from formative (paper-based) assessment in the classroom and tailor make lessons and revision on such specific analysis, so to achieve that feedback at the click of a mouse would save an enormous amount of time and labour, thus avoiding repetition.

In the short term, the many varied CAA and e-learning resources, like Mathletics, will serve us well as an evolution of diagnostic tool available to students and teachers.

9.3: A classification of student errors leading to improving CAA questions

- Can we infer, identify and classify student mistakes using evidence from students' work (ResultsPlus)?
- Can we generate a taxonomy for student errors?
- Can we use these known student mistakes and our own evidence basis (previous chapters) to improve CAA?

The data from the ResultsPlus files, while informative to the teacher, and quite informative to the student, really doesn't give the student many further practice materials or skill or topic specific questions that they can go away and practise repeatedly. ResultsPlus is also not as informative as we would hope in categorising student errors as our pilot paper-based (and onscreen) test was. Evidence of students' mistakes is readily inferred from results in chapter 3, and equally easily classified using the SOLO taxonomy in chapter 4. It is once we have successfully classified them that we can start to improve CAA questions by improving the mal-rules and hence distracters in the options.

In conclusion from the work completed in chapters 7 and 8, our aims when constructing improved CAA questions should be:

- Well-defined pedagogy for the topics / skills (items). To increase the efficacy of these tests, the learning objective of each of the item is determined before the construction of the questions.

- Parity across the questions on similar topics, so that no student produces a lower assessment score because they have randomly been allocated the one anomalously difficult question on that topic owing to structure, choice of distracters or fiddly nature of the numerical inputs.

- A careful evaluation of specific errors made by students in order to tailor-make their feedback. Such accurate diagnosis of errors and skills / topic improvements should lead to tailor-made questions easily accessible to students so that they can independently prepare for the summative assessment (whether for the first or subsequent time), especially when class time is often devoted to new skills / topics rather than the summative assessment.

9.4: The need for accurate selection of mal-rules

- Can we understand how the choice of mal-rules affects the difficulty of multiple-choice objective questions on calculus?
- How can we avoid commonality between distracters giving the answer away?
- Can we generate a reliable model for choosing and refining mal-rules for diagnostic calculus questions?

Students tend to be drawn to integer solutions or “nice numbers” e.g. ½ rather than more random decimal or surd answers, and such distracters can be easily discarded by students keen to take a short-cut. Lazy students (whether they are clever or not) will instinctively approach every question and systematically seek out the mal-rules that look obvious. On the other hand, the student who genuinely misunderstands the question may give a similar response by guesswork.

What the research undertaken in chapters 7 and 8 has particularly shown is that some objective multiple-choice questions have a high proportion of: “None of these” wrongly selected, and in some questions, a disproportionately high choice of a particular distracter over the other possible distracters. In questions where the majority of student errors were of “None of these”, tells us most likely that none of the other options worked well as a distracter, or that the student guessed. They also add little to our understanding of common student mistakes, and the MC objective question becomes more of a NI question instead.

The level and choice of distracters is also very clearly linked to the difficulty of the questions, so it is worth revisiting mal-rules often to make sure the questions are not overly easy or difficult relative to other similarly pedagogical questions owing to our choice of mal-rules.

However we structure mal-rules to build our distracters, we need to encourage the students to work through the questions carefully without guessing. The very nature of writing suitable responses for MCQs is key to completing the testing aims of these questions. My main conclusions for writing such responses are:

- Select distracters that are not too similar to each other (commonality), as this can lead to selection by elimination.
- Make sure we are meeting our objective test aims so that the correct skills are still being tested, and not fitting our responses to specifically hide the correct answer.
- Be aware that fiddly multi-structural problems that either require RNI or a selection from a MCQ list have very carefully chosen mal-rules that are completely evidence based mal-rules, not theoretical mal-rules, so that students don’t select “None of these” too often.
- Be aware that questions with fractional / decimal inputs as part of the multi-structural approach are perceived to yield fractional / decimal responses, so distracters in-built to produce integer responses are more likely to be intuitively eliminated.

- The facility of a question changes greatly when there is no guidance or strategy given to the student (no explicit cues), so a pedagogically identical set of responses to a question can give very different responses depending on how the question stem is set up.
- Students won't use "Don't know" as a response, but will use "None of these" often when unsure, and as such the response yields no useful information about their pre-structural mistakes.

9.5: The need for an accurate classification of mal-rules

Given the over-arching finding of this thesis is the importance of writing carefully constructed mal-rules for distracters of objective test questions. It is also worth looking carefully at how we classify them, so to better understand how we improve our selection policies. **Baruah (2007**) defined mal-rules as such: "for each distracter, the associated mal-rule is defined to be understood as the learning weakness(s) behind such item."

These mal-rules are determined from the structure of the problem (break points – where an error could be made or is likely to be made) or evidence *a posteriori* by administering the tests and analysing the answer files or from exam scripts.

By breaking this determination down into three classes, I conclude that there are times when each class is appropriate, and when some classes are the most appropriate for determining appropriate mal-rules. They can be categorised as follows:

- Evidence-based mal-rules

Genuine mal-rules as seen through paper-based test solutions, feedback from exam board examiners' reports, observed errors in marking random centre exam scripts (chapter 3), and confirmed errors selected by students using MC or RNI objective tests onscreen.

- Theory-led mal-rules

More anecdotal derived mal-rules, ideas generated by student questions and errors on paper in the classroom, recent surveys of perceptions of mathematics teachers and lecturers, mistakes generated by students when purposely trying to derive mal-rules themselves (see section 9.7). Idea for cases where there is no empirical data as to common student errors.

- Solution-led mal-rules

Mal-rules derived by splitting a multi-structural problem into its uni-structural processes and isolating a common known mistake at each independent process, while then following through to a solution, e.g.:

$$\int x^p dx = \frac{x^{p+1}}{p+1} + c$$

Here is power *p* is the fraction 1/10, so

1/(*p*+1)= 1/(1/10 + 1)= 10/(1+10). Consequently:

$$\int -3x^{1/10} dx = \frac{10 \times -3}{1+10} x^{(1+10)/10} + c$$

which, after some arithmetic, gives the answer.

Figure: 9.1: The worked multi-structural solution to an integration question

In every line, there is scope for making a mistake, even if following the correct general theory. The student could easily use (p-1) instead of (p+1) for both or either the x-power or the divisor. They could equally fail to correctly evaluate the divisor if it turns out to be a fraction (i.e. mis-remembering that dividing by 1 ½ is the same as multiplying by 2/3), and they could easily fail to correctly multiply the constant in front of the x term by the new (1/divisor).

The likely solution-led mal-rules for re-coding as distracters for such topic questions:

1) Not remembering integration rules, i.e. that the power of x goes up by 1 to p+1
2) Not remembering that the new divisor is p+1 in the integrand
3) Dividing by a fractional (p+1) is the same as multiplying by a fractional (p+1)
4) Multiplying by the negative of the divisor -(p+1)
5) Forgetting the + c for the indefinite integral
6) Differentiating the term correctly

This is not an evidence based collection of mal-rules, as there is no evidence that students would back up these choices, but if we assume that students have a rudimentary understanding of the topic and skill, then they are likely to make uni-structural mistakes, so a systematic breakdown of every possible opportunity to go wrong on the multi-structural route should lead us to a number of (student) easy to achieve mal-rules. How often these mistakes did indeed crop up would need further study.

9.6: Further conclusions and observations from the results

We can build really good objective tests year on year by using the previous year's results. Not only should we build up mal-rules again each year to take into account the mistakes the previous year's cohort had arrived at, but we should also look carefully at a selective exam based on those questions that yielded significant discrimination (above 0.5) in two successive years.

This work doesn’t have to be left entirely to the mathematical domain. I postulated in chapter 4 that the pedagogy of learning a modern foreign language was equally built on Bloom’s taxonomy, or indeed more up-to-date versions such as SOLO, and the remit to explore the use of diagnostic testing using similar objective tests for learning languages is very apparent to me. It could equally be applied to musical training, as the rationale for systematic elimination of errors is common here also.

9.7: Recommendations for future work

From the current study, it is clear that e-assessment is successfully implemented in tertiary education, at least at the early diagnostic stage of the courses, and is being piloted across schools at GCSE level. We do expect that students with A Level maths would do well in these Calculus and Algebra tests, but anecdotal evidence suggests that many students arrive at university “trained” to achieve a grade C or B, yet have very limited or transferable understanding. This would in itself be an ideal starting point for researching to what extent a notional grade at A Level justified certain assumed learnt skills.

There are many other research questions that have come to light during the course of this work that I attach here with brief discussion of where to start;

- How can CAA / e-assessment be embedded in secondary schools curricula?

Following discussions with a senior manager at the Edexcel exam board, I have been led to believe that we may see the online GCSE being the only mode of summative assessment within 10 years. Thus the many varied constraints on computer suites in schools, network capabilities, and practice resources will need to be planned now for such assessments in the next decade. What is clear is that these assessments will be going across to online versions, and so it isn’t an option of whether schools can embed e-assessment into their curricula but when they do it. After all, there are many benefits and opportunities for them to take advantage of once they go down this road.

- What are students’ own perceptions of mal-rules? Can they identify likely errors themselves when learning topics? Will the exercise help them to improve understanding and future performance?

As an experiment, I put a group of year 9, 10, 11 and 12 students through an paper-based test exercise that was appropriate to their levels, and asked them to find not only the correct solution but also identify three wrong solutions to go alongside the correct one. The year 9 students, on basic GCSE maths, found the correct solution relatively easily enough, but their choice of distracters was almost impossible to discern, as they seemed to make up responses, and I could hypothesise that they didn’t fully understand what mistakes other students could make as they only saw either the correct approach or had no clue how to solve the questions. The year 10 students were set an exercise on GCSE statistics, and their responses were more mature, yet often peppered with either random or superfluous responses that didn’t yield any generic patterns. Yet for the 29 year 11 and 12 students who sat an identical exercise on basic C1 level mathematics, their solutions yielded many of the same natural mistakes as in table 4.6. They perceived many natural uni-structural mistakes such as common sign errors, failure to complete the multiplication or subtract one from the power when differentiating a polynomial function, or forgetting the “+C” with an indefinite integration question. Some even spotted pre-structural mistakes such as differentiation and integration swapped around in a question.

Clearly this isn’t a scientific study as they did receive verification from me during the exercise and were allowed to confer with one another, but it does add useful anecdotal evidence supporting my findings in chapter 3 and 4, and should encourage others to conduct a larger survey using students to diagnose potential mal-rules. The learning benefits are three fold; namely that they practise useful formative assessment questions, they gain an understanding of likely mistakes so have a greater chance of remembering not to make them in the future, and also benefit assessment authors with further accurate mal-rules to incorporate. After all, the Duckworth-Lewis algorithm for determining rain-affected cricket matches is an evolving algorithm that takes into account every match played in history. In the youngest format of cricket (20 over innings), the early matches provided little evidence for which to formulate the model, but subsequent iterations with ever more data have proved to be much more accurate, so the same should hold here with more evidence of student errors for which to base our mal-rules on.

- What use do students actually make of feedback and how effective / ineffective is it?

Also see page 153 to further explore the measurable benefits of CAA testing. As discussed previously, the efficacy of post question feedback, tailor-made support, whether through e-learning, maths clinics, additional teaching time, one-to-one tutor support or else is something that needs investigating further. Anecdotal evidence suggests that students do look at feedback, but this has come from various unscientific hand-raising surveys in lecture halls. If the software for e-assessment is useful, then we need to investigate how we collect evidence that students learn anything from it (e.g. questionnaires and analysis of exam scripts).

- What would be the effect of mixed question types, different scoring systems or use of confidence based questions?

We have plenty of options for development of different question styles, e.g. more RNI, Hot Line, or Sequential questions using Mathletics and many have already been written for calculus topics already. The challenge is to make them objective enough to deliver useful feedback on mistakes made, and robust enough to be sustainable through multiple usages. I also ask whether different modes of scoring on questions would be worth exploring. Some work has been done on confidence based questions (where students give a proportion of their confidence in their own answer, or select from a number of word ranges of confidence), and thus the higher the confidence the more marks they gain if correct and equally more lost if wrong. Different scoring systems (like confidence based questions, or negative marks for wrong answers) force students to think strategically, and may well counter a guessing culture amongst them. The Russian scoring system (http://www.mathcomp.leeds.ac.uk/pdfs/IMOK%20guidance%20notes.pdf) dictates whether a question is substantially correct (scoring between 70-100%) or substantially wrong (scoring 0-40%). Such a model will complicate the marking as the negative scoring will look punitative, and the positive marking will look generous.

- How robust are the objective questions for onscreen GCSEs? Especially how are the options for students determined? Do they fully test the same skills as paper-based GCSEs?

The example of one practice GCSE question shown in chapter 2, shows a robust looking question with no distracters that looked immediately wrong, so wasn't intuitively likely to lead to a guess from the students. It remains to be seen whether online GCSE practice questions will ever be robust and reusable, as the aim of the GCSE summative assessment is to set novel questions testing the same skills. A piece of research on these tests as they become main-stream would also answer the questions of how the options are determined and whether the same skills are tested exactly as paper-based tests, although again it remains to be seen whether specific mal-rules are built in to options to offer such specific feedback (as opposed to just generic feedback including a score, topic analysis and worked solutions).

- Can we generate a multi-use diagnostic test using excel that has a web-style user interface and provides onscreen feedback as easily and quickly as any other diagnostic tool used in e-assessment?

As shown earlier, the format of the Excel worksheet makes it very difficult to clearly depict anything other than basic arithmetic, as typesetting is done by columns and some fractions look very odd when they arrive onscreen. I also suggested that those with expertise in Visual Basic could write a macro that gives a number of selection boxes for multiple-choice answers for the questions, and that a final submission box would lead to a revised screen with analysis of the number of correct responses and feedback on which errors were likely to have been made with any wrong selections. Such feedback could record the marks for that student as a summative assessment, but would be much more suitable to formative assessment. However, while random parameters make such tests reusable, we are stuck with the same style of question realisations, rather than the much better ability to pick a selection of questions from a database – a series of diagnostic tests spanning skills is much more useful, just as a specific results profile is more important than the marks.

We did, through results in chapters 6 and 7, generate a very hard C1 practice paper which could equally be turned into a multi-use Excel worksheet and also some specifically hard calculus questions for EC1005 students that could also be turned into a multi-use excel spreadsheet if we wanted to stretch the best students or offer warnings to those who we perceived weren't working hard enough.

- Can we compile a full dictionary of student errors or mistakes by topic or skill level using evidence from others?

There is more than a large thesis of work that could be undertaken here, and a plethora of common error sources that could be tapped (see table 9.2), to better guide us with encoding mal-rules / improving taxonomy of errors.

The first 8 sources below would serve as a good starting point for such work, half from the US and half from the UK. A spin-off question would look at the differences (if any) between errors derived from the US high-school teaching world and that in the UK.

www.math.vanderblit.edu/~schechtex/commerrs/	US
http://mathmistakes.org/	US
www.calvin.edu/~schofield/courses/materials/tae/	US
www.teachernet.gov.uk/teachers/issue42/rimary/features/Mathsmisconceptions/	UK
www.math.hmc.edu/calculus/tutorials/algebrareview/	US
www.mathsyear2000.co.uk/resources/misconceptions/	UK
www.cimt/plymouth.ac.uk/resources/help/miscon.htm	UK
www.amazon.co.uk/gp/product/1844450325/ref=sib_rdr_dp	UK

Table 9.2: A resource list of generic mistakes made by students in mathematics.

- Can we objectively demonstrate the benefits of CAA testing and measure the performance effects?

Again, this question arises as a consequence of seeing an increase in pass marks on paper-based summative assessments (A Levels etc year on year), and also the improvement in scores on the summative tests for EC1005 students. **Greenhow (2009)**, pointed out that exam marks for EC1005 increased by 5% by from the 2007-2008 to 2009-2009, with the base intake analysis of the cohort the same for both years. Within chapter 7, we conclude also that general performances rise by 3% from 2008-2009 to 2009-2010. He hypothesised that this was down to the formative capability of Mathletics and use of Maths café by students during the semester. Clearly much careful control samples and temporal analysis over a number of cohorts would need to take place before generalisations could be made.

- Will we generate significantly different results if we looked at mistakes made solely by a sample from a school / college that were made up from a 100% female cohort? Or a comprehensive school with a mixed cohort?

Warwick and Harrow schools are independent secondary schools, both with male only cohorts. The nature of selection from these institutions is primarily for ease of access of a large quantity of sample data, and this may well yield less certain generalisations of a parent population. The results taken from my own marking of old style P1 and P2 papers drew from a random set of centres, so cross educational establishments cross gender and provided some very useful backup of earlier findings. A greater selection of centres would, indeed, provide a rich set of data in which to compare these findings.

- What can we glean from the nature of retakes of module or unit papers? Will the same mistakes be generated on the retake or will they be fewer of the same mistakes, or yet more different mistakes?

Of the 65 retake papers in C1 sat by our students, all but 9 of them either improved their % on differentiation or achieved the same (100%) understanding of the topic 2^{nd} time round. The average improvement was 12% on this topic alone from the 1^{st} to the 2^{nd} retake. However, for those who sat the paper for a 3^{rd} time, 4 improved, 8 stayed the same and 5 went down (an average of 1 %). Clearly this is a topic where the weaker candidates are learning to some extent from their mistakes, but the very weakest candidates are not learning enough from their returned exam scripts or revision work to make the inroads on this particular topic. A computer-based learning and assessment scheme is therefore likely to yield the sort of improvement that they need by the 3^{rd} attempt at such a paper.

References

Anson, C. (2000). 'Response and the social construction of error'. Assessing Writing, 7(1), p5-21.

Baker, M. (2010). 'Are exams really getting easier?'. www.bbc.co.uk/news/10094248

Barry, M. et al. (2003). 'Diagnostic Testing and Student-Centred Support'. Proceedings: Mathematical Education of Engineers, IMA.

Baruah, N. (2007). 'Computer-aided assessment of higher-level calculus: design and evaluation'. Brunel University, PhD Thesis.

Biggs, J. (1999). 'Teaching for Quality Learning at University'. SRHE and Open University Press, Buckingham

Biggs, J. & Collis, K. (1982). 'Evaluating the quality of learning: The SOLO Taxonomy.' New York. Academic Press.

Black, P & Wiliam, D (2010). 'Developing the theory of formative assessment. Educational Assessment'. Evaluation and Accountability, 21(1), p5-31.

Bloom, B. (1956) Taxonomy of Educational Objectives, the classification of educational goals – Handbook I: Cognitive Domain New York: McKay

CAA (2002). Computer-aided assessment. www.caacentre.ac.uk

Carson & Dale (1909). University of London, Schools Matriculation Exam (www.edexcel.com/emporium).

Chick, H. (1998). ‘Cognition in the formal modes: Research Mathematics and the SOLO Taxonomy’. Mathematics Education Research Journal, Vol 10, No.2 4-26.

CIE (2009). ‘eAssessment in action – Geography’, Cambridge International Examinations group, www.cie.org.uk

Clarke, S (2010). ‘Unlocking Formative Assessment - Practical strategies for enhancing pupils' learning.’ http://www.highlandschools-virtualib.org.uk/ltt/inclusive_enjoyable/formative.htm

Crocker, L. & Algina, J. (1986). ‘An introduction to classical and modern test theory’. New York: Holt, Rinehart and Winston.

Cumming, G (2010). ‘ResultsPlus demonstration’, Edexcel Examinations Group. www.edexcel.com

Dalton, T (2006). Queen Elizabeth School for boys, Department of Mathematics, London. - personal communication

Davies, G. (2010). ‘Practice: the missing link’. http://ictforlanguageteachers.blogspot.com/2010/02/practice-missing-link.html.

Drought, P. (2010). Edexcel Examinations Group, personal communication.

Earl, S., Land, R. & Wise, J. (2000). ‘Assessment, Evaluation and Support.’ MSc in Computer-enhanced Mathematical Education, Napier University.

Ellis, N. (2010). In ‘Pressure high as students await GCSE results’. www.bbc.co.uk/news/education-11064110

Emanuel, R. (2010). ‘School mathematics exams – a view from a head of department’. Mathematics Today, IMA, Vol 46, p133.

Galton (1888); in Venn, J. & J. A., Alumni Cambridgiensis, Cambridge University press.

Gibson, I. (2004). University of Glasgow, Mathematics Diagnostic tests, www.gla.ac.uk/maths

Gill, M. & Greenhow, M. (2006). 'Computer-aided assessment in Mechanics: question design and test evaluation'. Teaching Mathematics and its Application, MEE.

Greenhow, M. (2010). – Brunel University, department of mathematics - communication of results files

Greenhow, M. (2009). 'Mathletics – a suite of computer-aided assessments'. Brunel University, guidance for use.

Greenhow, M. (2009). – Brunel University, department of mathematics - communication of results files

Greenhow, M. et al. (2003). 'Pedagogic issues in setting online questions'. MSOR Conns 2(3): August, p19-22. http://ltsn.mathstore.ac.uk/headocs/34pedagogic.pdf

Greenhow, M. (2000). 'Setting objective tests in mathematics with QM designer'. MSOR, LTSN: http://ltsn.mathstore.ac.uk/newsletter/feb2000/pdf/qmdesign.pdf.

Greenhow, M. et al. (1999). Mathletics – a suite of computer-aided assessments, Brunel University 1999 -.

Greenhow, M. (1996). 'Computer based diagnostic tests in mathematics for QM designer'. Quarterly newsletter of CTI Maths and Stats 7 (3) Aug 1996 20-24.

Golding, J. (2010). ‘School mathematics exams – a view from a chartered mathematics teacher’. Mathematics Today, IMA, Vol 46, p132.

Harper, R. (2003). ‘Multiple-choice questions – a reprieve’, BEE-j Vol 2. http://bio.ltsn.ac.uk/journal/vol2/beej-2-6.htm.

Harrison, A. (2010). ‘A-Level results 2010: A* grade boosts new exams record’. http://www.bbc.co.uk/news/education-11012369.

Harte, C. (2010). ‘Going SOLO – assessing understanding of language concepts’. Learner Evolution. http://chrisharte.typepad.com/learner_evolution_chris_h/2010/

Hatt, J. & Baruah, N. (2006). ‘The Reconceptualisation of the revised Bloom’s taxonomy for use in Mathematics and its Implementation into QM Perception and Mathletics’. PRHE Conference, Liverpool, May. http://hopelive.hope.ac.uk/PRHE/

IMA (2010). ‘Keep maths textbooks independent of examining boards’. http://www.ima.org.uk/Education/SFE.htm.

IMA (2003). Institute of Mathematics and its Applications, Spring review, www.ima.org.uk.

Jones, M. (2010). ‘GCSE Revision day at the University of Glamorgan’. Mathematics Today. www.ima.org.uk

Lawson, D. (2003). ‘Lessons from a decade of diagnostic testing’. Proceedings: Mathematical Education of Engineers, IMA.

Lawson, D. (2002). ‘Computer-aided assessment in mathematics: Panacea or propaganda?’. CAL-laborate, Oct. 2002, p6-12, UniServe Science, Sydney University

Lomax, R. (2000). ‘An introduction to statistical concept for Education and the Behavioural’. Lawrence Erlbaum Associates Inc, US, ISBM: 0805827498.

Mann, H & Whitney, D (1947). ‘On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other’. Annals of Mathematical Statistics 18, p50-60.

Marzano, R. (2000). ‘Designing a new taxonomy of educational objectives’. Thousand Oaks, CA: Corwin Press.

Massey, A. (1995). ‘Evaluation and analysis of examination data: Some guidelines for reporting and interpretation’. UCLES internal report, Cambridge.

McAlpine, M. (2002). ‘Principles of assessment’. Blue paper No. 1 Bedfordshire: CAA centre, University of Luton.

McKenna, C. and Bull, J. (2000). ‘Quality Assurance of Computer-assisted Assessment: Practical and Strategic Issues’, Quality Assurance In Higher Education. Vol. 8, No. 1. p. 24-31.

MEI (2010). www.mei.org.uk

Murphy, R. et al. (2010). ‘School exams: have standards really fallen?’. Nottingham University at Cambridge Assessment seminar. www.cambridgeassessment.org.uk

Nicol, D. & McFarlane-Dick, D. (2004). ‘Formative assessment and self-regulated learning: a model and seven principles of good feedback practice’. Studies in Higher Education, Vol 31(2), p111-218.

Orton, A. (1983). ‘Students’ understanding of integration’. Educational Studies in Mathematics, 14, p1-18.

Paterson, J. (2002). 'What's in name? – A new hierarchy for question types'. Proc. 6th Intl CAA conference, Loughborough University, 2002, 285-292.

Pearson, K. (1917). 'On Lines and Planes of Closest Fit to Systems of Points is Space'. Philosophical Magazine Series 6 2: 559–572

Spearman, C. (1904). 'General intelligence objectively determined and measured.' American Journal of Psychology, 15, 201–293

Priem, J. (2010). 'Fail better: Toward a Taxonomy of e-learning error'. J.Educational Computing Research, Vol. 43(3) pp377-397.

Puttnam, D. (2011). 'Digital Curriculum'. Chairman of Futurelab; www.mrstucke.com/2011/

Ricketts, P. (2010). 'Questions of value: A-Levels, schools and higher education'. The Times, August 17 2010, Letters to the Editor. www.thetimes.co.uk

Rider, R. (2010). 'Using the SOLO Taxonomy to evaluate learning of function concepts in developmental algebra'. East Carolina University.

Robinson, C & Croft, A. (2003). 'Diagnostic testing and support for engineering students', Proceedings: Mathematical Education of Engineers, IMA.

Ross, J. (2011). 'E-Assessment Question'. www.e-assessment-question.co.uk.

Rowlett, P. (2010). 'E-Learning in Mathematical Subjects', Mathematics Today.

Schechter, E. (1994). 'The most common errors in undergraduate mathematics'. http://www.math.vanderbilt.edu/~schectex/commerrs/

Smithers, A. (2010). ‘School exams: have standards really fallen?’. Buckingham University at Cambridge Assessment seminar. www.cambridgeassessment.org.uk

Spriggs, M. (2010). ‘Questions of value: A-Levels, schools and higher education’. The Times, August 17 2010, Letters to the Editor. www.thetimes.co.uk

Todd, P. (2002). ‘The correlation between A-level maths grades and subsequent performance’. Proceedings: Mathematical Education of Engineers, IMA.

Watson, J. (1994). ‘A diagrammatic representation for studying problem-solving behaviour.’ J. Mathematical Behavior, 13, pp305-332.

Watson, J., Chick, H. & Collis, K. (1988). ‘Applying the SOLO Taxonomy to errors on area problems’. J.Pegg (Ed.) Mathematical Interfaces (Proc 12th biennial conference of the Australian Association of Mathematics Teachers, pp. 260-281).

Webster, M. (1985). ‘Webster`s ninth new collegiate dictionary’. Meriam - Webster Inc.

Wood, D. (1960). ‘Test construction: Development and interpretation of achievement tests’. Columbus, OH: Charles E. Merrill books, inc.

Wood, P (2009). E-assessment feedback via CIE, New English School, Kuwait.

Appendix 1: C1 pilot diagnostic test with multiple-choice distractors built in:

Question	Question	Multiple choice solutions						Answer
		A)	B)	C)	D)	E)	F)	
1	Evaluate: $\lvert -26+10 \rvert$	-16	|16|	36	16	-36	None	
2	Evaluate: $\frac{-3}{5}\left[\frac{2}{3}-\frac{1}{2}\left(\frac{3}{4}-\frac{1}{2}\right)\right]$	-1/40	-13/40	-3/120	-1/16	-4/72	-39/120	
3	Factor: $15x-3$	x = 1/5	3(5x-3)	3(5x-1)	5(3x-1)	5x-1	x = 3/15	
4	Simplify: $\frac{5}{x}-\frac{7}{y}$	$\frac{(5y-7x)}{xy}$	5y-7x	$\frac{-2}{(x-y)}$	7x-5y	$\frac{(7x-5y)}{xy}$	None	
5	Simplify: $5x-[2x+4(x-3)]$	7x-12	$2x^2$+x-12	3x-8	30(x^2-2x)	12-x	None	
6	Solve for *x*: $\lvert x+4 \rvert = 6$	-2	2, 10	2	2, -2	2, -10	None	
7	Simplify: $\frac{x^2-4}{3x}\cdot\frac{9}{3x-6}$	x^2-6x+8	$\frac{-36}{-18x}$	$\frac{-2}{x}$	$3x^2$-18x+24	$\frac{x}{x+2}$	None	
8	Simplify: $\frac{4x-8}{x^2-x-2}$	$\frac{4}{x+2}$	$\frac{4x}{x+1}$	$\frac{4}{x+1}$	-4x-8	$\frac{-2}{x^2}$	None	
9	Simplify: $(-6a^2b^3)(2a^4b^2)$	$-12a^8b^5$	$-12a^8b^6$	$-12a^8b^8$	$-12a^6b^6$	$-12a^5b^5$	None	
10	Simplify: $\left(\frac{2x^7y^2}{4xy^3}\right)^3$	$\left(\frac{x^6}{2y^{-1}}\right)^3$	$\left(\frac{x^{16}}{8y^3}\right)$	$\left(\frac{x^7}{8y^3}\right)$	$\left(\frac{x^{18}}{8y^6}\right)$	$\left(\frac{x^{18}}{8y^3}\right)$	None of these	

11	Simplify: $(3x-5)(2x+7)$	A) 6x^2+11x-35	B) 6x^2-11x-35	C) 6x^2+11x+35	D) 5x^2+11x-35	E) 6x-35	F) None	
12	Solve for *x* $-6(2x+1)=-4x+10$,	A) -0.5	B) -1	C) -2	D) -0.25	E) -9/8	F) None	
13	Solve for *x*:: $2x-6\le -9+7x$	A) $3\le 5x$	B) $0.6\le x$	C) $3/5\ge x$	D) $-x\le 3$	E) $5/3\le x$	F) None	
14	Solve for *x*:: $\frac{2x}{5}+9=x+\frac{1}{6}$	A) $-15\frac{5}{8}$	B) 15	C) 11	D) $-14\frac{13}{18}$	E) $-8\frac{1}{6}$	F) None	
15	Solve for *x*:: $x^2-6x+5=0$	A) 1, 2	B) 1, 5	C) -1, -5	D) 1, -5	E) 1, -2	F) None	
16	If 7 is subtracted from three times a number, we get the same result as we get from adding 9 to one third of the number. What is the number?	A) 24	B) 4.8	C) 17	D) 6	E) 3	F) None	

Appendix 2: Sample feedback screen for Q 9-13 (answers A-C) of AS diagnostic test.

	A	B	C
9	$-24a^{20}b^{7}$	$-24a^{20}b^{12}$	$-24a^{20}b^{15}$
	Multiplied wrong first indices	Multiplied Indices Not added them	Multiplied wrong pairs and not added
10	$\left(\dfrac{x^{4}}{4y^{5}}\right)^{3}$	$\left(\dfrac{X^{36}}{64Y^{15}}\right)$	$\left(\dfrac{x^{4}}{64y^{15}}\right)$
	Not multiplied out brackets	Multiplied x powers not added	Not raised the x power
11	$28x^2 + 30x - 18$	$28x^2 - 30x - 18$	$28x^2 + 30x + 18$
	Correct	Wrong sign for x term	Wrong sign for x term
12	x = 6	x = 3	x = 12
	Factor of two out check division	Forgot to divide by denominator at end	Correct
13	$5 \leq 5x$	$x \leq 1$	$-1 \quad x$
	Not finished off In terms of x	Correct	Wrong sign on division

Appendix 3: C1 hard questions derived from ResultsPlus.

Paper Reference(s)

6663/01

Edexcel GCE

Core Mathematics C1

Advanced Subsidiary

Hard C1 past questions

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Pink or Green)	Nil

Calculators may NOT be used in this examination.

Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Core Mathematics C1), the paper reference (6663), your surname, initials and signature.

Information for Candidates

A booklet ‘Mathematical Formulae and Statistical Tables’ is provided.
Full marks may be obtained for answers to ALL questions.
The marks for the parts of questions are shown in round brackets, e.g. **(2)**.
There are 10 questions in this question paper. The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

1.

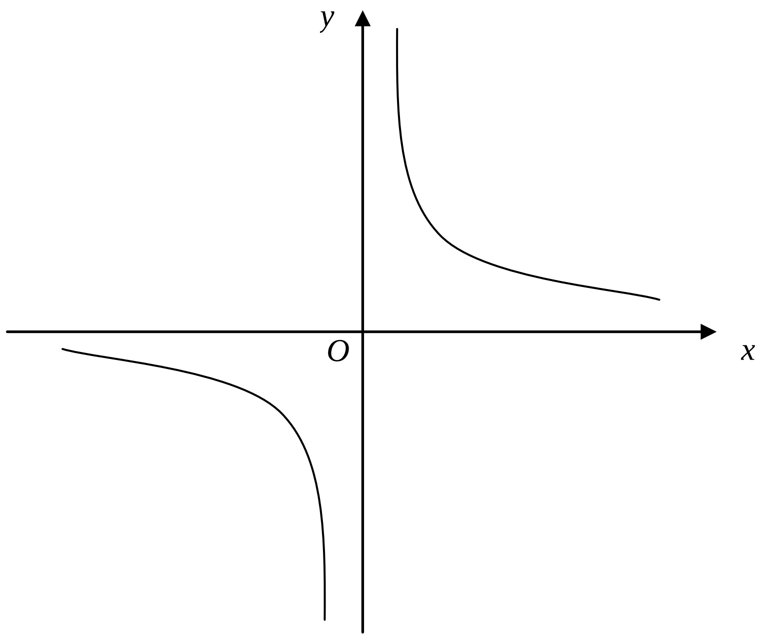


Figure 1

Figure 1 shows a sketch of the curve with equation $y = \frac{3}{x}, \quad x \neq 0$.

(*a*) On a separate diagram, sketch the curve with equation $y = \frac{3}{x+2}, \quad x \neq -2$, showing the coordinates of any point at which the curve crosses a coordinate axis.

(3)

(*b*) Write down the equations of the asymptotes of the curve in part (*a*).

(2)

---

2. (*a*) By eliminating *y* from the equations

$$y = x - 4,$$

$$2x^2 - xy = 8,$$

show that

$$x^2 + 4x - 8 = 0.$$

(2)

(*b*) Hence, or otherwise, solve the simultaneous equations

$$y = x - 4,$$

$$2x^2 - xy = 8,$$

giving your answers in the form $a \pm b\sqrt{3}$, where a and b are integers.

(5)

---

3. The equation

$$x^2 + kx + 8 = k$$

has no real solutions for x.

(a) Show that k satisfies $k^2 + 4k - 32 < 0$.

(3)

(b) Hence find the set of possible values of k.

(4)

---

4. The first term of an arithmetic sequence is 30 and the common difference is -1.5.

(a) Find the value of the 25th term.

(2)

The rth term of the sequence is 0.

(b) Find the value of r.

(2)

The sum of the first n terms of the sequence is S_n.

(c) Find the largest positive value of S_n.

(3)

---

5. Simplify

(a) $(3\sqrt{7})^2$

(1)

(b) $(8 + \sqrt{5})(2 - \sqrt{5})$

(3)

---

6. Given that $32\sqrt{2} = 2^a$, find the value of a.

(3)

---

7. Given that $y = 2x^3 + \dfrac{3}{x^2}$, $x \neq 0$, find

$\dfrac{dy}{dx}$,

---

(3)

8.

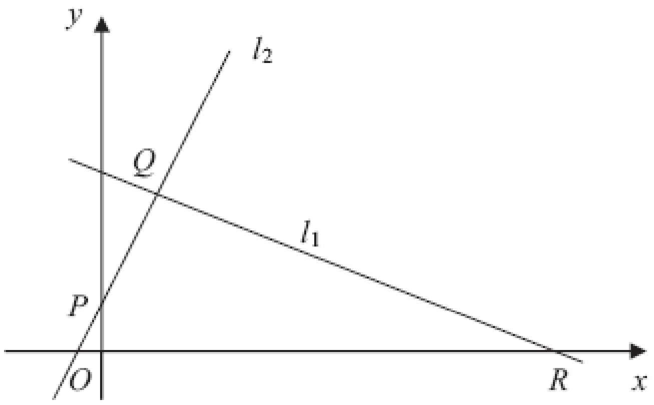


Figure 2

The points Q (1, 3) and R (7, 0) lie on the line l_1, as shown in Figure 2.

The length of QR is $a\sqrt{5}$.

(*a*) Find the value of *a*.

(3)

The line l_2 is perpendicular to l_1, passes through Q and crosses the y-axis at the point P, as shown in Figure 2. Find

(*b*) an equation for l_2,

(5)

(*c*) the coordinates of *P*,

(1)

(*d*) the area of ΔPQR.

(4)

---

9. The gradient of a curve C is given by $\dfrac{dy}{dx} = \dfrac{(x^2+3)^2}{x^2}$, $x \neq 0$.

The point (3, 20) lies on C.

Find an equation for the curve C in the form $y = f(x)$. **(6)**

---

10. The line l_1 passes through the point $A(2, 5)$ and has gradient $-\frac{1}{2}$.

(a) Find an equation of l_1, giving your answer in the form $y = mx + c$.

(3)

The point B has coordinates $(–2, 7)$, and lies on l_1.

(b) Find the length of AB, giving your answer in the form $k\sqrt{5}$, where k is an integer.

(3)

The point C lies on l_1 and has x-coordinate equal to p.

The length of AC is 5 units.

(c) Show that p satisfies

$$p^2 - 4p - 16 = 0.$$

(4)

---

11. Given that $y = x^4 + x^{\frac{1}{3}} + 3$, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

(3)

---

12. $$\frac{\mathrm{d}y}{\mathrm{d}x} = 5x^{-\frac{1}{2}} + x\sqrt{x}, \qquad x > 0.$$

Given that $y = 35$ at $x = 4$, find y in terms of x, giving each term in its simplest form.

(7)

Appendix 4: Screenshot of C1 practice paper (dynamic) and answers (Excel).

C1 practice paper (dynamic)

Q1 Let $a_1 = k$

and $a_{n+1} = 7a_n + 4$

Q1a Find a_2

Q1b Show that $a_3 = 49k + 32$

Q1c Work out (in terms of k) $\sum_{r=1}^{r=4} a_r$

Q2a Find: $\int 7x - \frac{6}{x^2}\, dx$

Q2b Given that $y = 4x - \frac{12}{x^2}$

Find: $\frac{dy}{dx}$

Q3 Given that $y + 4x = -1$

and $2x^2 - 13x - y = 19$

Solve to find x and y

Q4

Q4a Solve: $5(6x + 5) < 7 - 28x$

Q4b Solve: $x^2 - 18x + 36 > 0$

Q4c For which region is: $5(6x + 5) > 7 - 28x$

and $x^2 - 18x + 36 > 0$??

C1 practice answers (dynamic)

1 Using the information given in both equations:

$a_1 = k$

1a $a_2 = 7a_1 + 4 =$

1b $a_3 = 7a_2 + 4 =$

1c $\sum_{r=1}^{r=4} a_r = a_1 + a_2 +$

$a_4 = 343k + 228$

$\sum_{r=1}^{r=4} a_r = 400k + 264$

2a $\int = 3\,1/2\,x^2 + \frac{6}{x} +$

2b $\frac{dy}{dx} = 4 + \frac{24}{x^3}$

3 Subbing the top equation into the bottom one: $y =$

$2x^2 - 13x - y =$

$2x^2 - 13x - (-1$

$2x^2 - 9x - 18 = 0$

$(2x + 3)(x - 6) =$

$x = -1\,1/2$ or 6

$y = 8$ or -37

4a $x < -9$

4b $x > 6$ or $x < 3$

4c $x < -9$

Appendix 5: Description of the questions included in differentiation tests for EC1005

Question description	Topic
sqrt((x^a)^b); MC	Differentiation\Algebraic functions
sqrt(x^a); MC	Differentiation\Algebraic functions
(ax^b+c)^n,x,a,b,c); MC	Differentiation\Chain rule\Binomials
(ax+b)^3; HL	Differentiation\Chain rule\Binomials
(ax+b)^n; a,b +ve/-ve,n+ve; NI	Differentiation\Chain rule\Binomials
(ax+b)^n; a,n +ve/-ve; RNI	Differentiation\Chain rule\Binomials
(ax+b)^n; a,n +ve; MC	Differentiation\Chain rule\Binomials
(b+cx)^3 ; b,c +ve; MC	Differentiation\Chain rule\Binomials
diff ax^b; a,b +ve; MC	Differentiation\Polynomials
diff cubic version 1; MC	Differentiation\Polynomials
diff cubic version 2; MC	Differentiation\Polynomials
diff cubic; +ve coeffs; evaluate at x=1/a; MC	Differentiation\Polynomials
diff cubic; +ve coeffs; evaluate at x=a; MC	Differentiation\Polynomials
diff polynomial n=4...12 version 1; MC	Differentiation\Polynomials
diff polynomial n=4...12; MC	Differentiation\Polynomials
diff polynomial n=4...6; +ve coeffs; evaluate at x=1/a; MC	Differentiation\Polynomials
diff polynomial n=4...6; +ve coeffs; evaluate at x=a; MC	Differentiation\Polynomials
diff quadratic version 1; MC	Differentiation\Polynomials
diff quadratic version 2; MC	Differentiation\Polynomials
diff quadratic; +ve coeffs; evaluate at x=1/a; MC	Differentiation\Polynomials
diff quadratic; +ve coeffs; evaluate at x=a; MC	Differentiation\Polynomials
Min of f(x)=Ax^2+Bx+C, A, B, C +ve; NI	Differentiation\Polynomials
x+y given. Find y for TP of xy^2. x,y +ve; NI	Differentiation\Polynomials
diff(cubic* Exp(cubic); MC	Differentiation\Product rule\Exponentials
diff(quadratic* Exp(ax); MC	Differentiation\Product rule\Exponentials
diff(quadratic* Exp(cubic); MC	Differentiation\Product rule\Exponentials
diff(quadratic* Exp(linear); MC	Differentiation\Product rule\Exponentials
diff(quartic* Exp(quadratic); MC	Differentiation\Product rule\Exponentials
diff cubic*cubic; MC	Differentiation\Product rule\Polynomials
diff quadratic*cubic; MC	Differentiation\Product rule\Polynomials
diff quadratic*quartic; MC	Differentiation\Product rule\Polynomials
diff quartic*quartic; MC	Differentiation\Product rule\Polynomials
diff(a+bx^2+cx^3)(dx+ex^3+fx^5); MC	Differentiation\Product rule\Polynomials

Appendix 6: Test on differentiation for EC1005 based on hardest questions

Such a test would include the following ten questions, where there has been no improvement:

Question descriptor	**Difference of means**
diff(quadratic* Exp(linear); MC	-0.062
(b+cx)^3 ; b,c +ve; MC	-0.053
diff quadratic version 1; MC	-0.037
(ax+b)^n; a,n +ve/-ve; RNI	-0.033
(ax+b)^3; HL	-0.03
diff(quartic* Exp(quadratic); MC	-0.025
diff polynomial n=4...6; +ve coeffs; evaluate at x=1/a; MC	-0.023
diff(quadratic* Exp(ax); MC	-0.021
sqrt(x^a); MC	-0.002
(ax^b+c)^n,x,a,b,c); MC	0

Appendix 7: Integration ASSESSMENT test report for EC1005 cohort 2008-2009.

Question description	Topic	Question type	Times presented	Times answered	Mean score	Difficulty	Outcome analysis			
(Ax+C)exp(bx); MC	Integration\By Parts\Exponential functions	Explanation	59	58	0.621	0.621	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	36	61.02%	88.89%
							integrated each term separately	6	10.17%	56.67%
							differentiated exp(Bx) instead of integrating	5	8.47%	56%
							incorrect sign in the formula	6	10.17%	50%
							differentiation instead of integration	0		-
							None Of These	5	8.47%	44%
							Did Not Know	0		-
							Not Answered	1	1.69%	40%
(x^2-Ax)e^x; NI	Integration\By Parts\Exponential functions	Explanation	72	72	0.694	0.347	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Both Correct	25	34.72%	88.60%
							None Right	47	65.28%	47.21%
							Not Answered	0		-
Axexp(Bx); MC	Integration\By Parts\Exponential functions	Explanation	63	62	0.758	0.758	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	47	74.60%	85.53%
							integrated each term separately	2	3.17%	30%
							differentiated exp(Bx) instead of integrating	4	6.35%	55%
							incorrect sign in the formula	2	3.17%	30%
							differentiation instead of integration	0		-
							None Of These	6	9.52%	36.67%
							Did Not Know	1	1.59%	40%
							Not Answered	1	1.59%	40%
int b x^(1/q); q +ve; MC	Integration\Algebraic functions	Explanation	232	231	0.887	0.887	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	205	88.36%	68.69%
							not p+1	3	1.29%	19%
							diff & not p+1	1	0.43%	0%
							diff	12	5.17%	32.25%
							guess	1	0.43%	0%
							None Of These	8	3.45%	19.13%
							did not know	1	0.43%	0%
							Not Answered	1	0.43%	40%
int b x^(1/q); q -ve; MC	Integration\Algebraic functions	Explanation	264	262	0.874	0.874	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	229	86.74%	65.47%
							not p+1	10	3.79%	28%
							diff & not p+1	5	1.89%	16%
							diff	5	1.89%	4%
							guess	7	2.65%	28.57%
							None Of These	6	2.27%	22.83%
							Did Not Know	0		-
							Not Answered	2	0.76%	60%
int b x^(p/q); q +ve; MC	Integration\Algebraic functions	Explanation	243	243	0.877	0.877	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	213	87.65%	67.12%
							not p+1	4	1.65%	32.50%
							diff & not p+1	2	0.82%	10%
							diff	14	5.76%	34.50%
							guess	4	1.65%	0%
							None Of These	5	2.06%	32%
							Did Not Know	1	0.41%	0%
							Not Answered	0		-
int((x/a)(exp(x/q+p)),x,a,q,p); MC	Integration\By Parts\Exponential functions	Explanation	75	75	0.507	0.507	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	38	50.67%	85.26%
							formula incorrect, sign.	8	10.67%	45%
							choose exp(x) to be v instead of dv	13	17.33%	44.62%
							integral of dv wrong, sign	9	12%	60%

							integral of dv wrong, constant	0		-
							None Of These	7	9.33%	68.57%
							Did Not Know	0		-
							Not Answered	0		-
int(a+bx+cx^2,x); MC	Integration\Polynomials	Explanation	89	88	0.886	0.886	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	78	87.64%	67.05%
							No Constant	0		-
							Integration (xply b+1)	2	2.25%	20%
							Differentiated	0		-
							Integration (no div b+1)	5	5.62%	26%
							None Of These	3	3.37%	13.33%
							Did Not Know	0		-
							Not Answered	1	1.12%	0%
int(ax(exp(px)),x,a,p); MC	Integration\By Parts\Exponential functions	Explanation	74	74	0.514	0.514	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	38	51.35%	83.68%
							formula incorrect, sign.	19	25.68%	51.58%
							choose exp(x) to be v instead of dv	6	8.11%	50%
							integral of dv wrong, sign	4	5.41%	35%
							integral of dv wrong, constant	0		-
							None Of These	6	8.11%	50%
							Did Not Know	1	1.35%	0%
							Not Answered	0		-
int(ax(exp(px/q)),x,a,p,q); MC	Integration\By Parts\Exponential functions	Explanation	73	72	0.528	0.528	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	38	52.05%	88.95%
							formula incorrect, sign.	15	20.55%	52%
							choose exp(x) to be v instead of dv	8	10.96%	45%
							integral of dv wrong, sign	2	2.74%	70%
							integral of dv wrong, constant	2	2.74%	50%
							None Of These	7	9.59%	42.86%
							Did Not Know	0		-
							Not Answered	1	1.37%	40%
int(ax^(n-1)/(bx^n+g),x); MC	Integration\Rational functions\Logarithmic form	Explanation	135	135	0.681	0.681	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	92	68.15%	82.11%
							Missing n	12	8.89%	46.67%
							Wrong Integration	15	11.11%	40%
							Substitute	4	2.96%	65%
							No Constant	3	2.22%	33.33%
							None Of These	7	5.19%	41.86%
							Did Not Know	2	1.48%	20%
							Not Answered	0		-
int(Ax^(n-1)/(Bx^n+G),x,a,b), a greater b +ve; RNI	Integration\Rational functions\Logarithmic form	Explanation	118	115	0.539	0.539	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	62	52.54%	87.60%
							dis1	0		-
							dis2	2	1.69%	40%
							dis3	1	0.85%	40%
							dis4	1	0.85%	40%
							dis5	0		-
							dis6	5	4.24%	36%
							unidentified error or Not Answered	44	37.29%	47.95%
int(Ax^(n-1)/(Bx^n+G),x,a,b), a greater b -ve; RNI	Integration\Rational functions\Logarithmic form	Explanation	138	133	0.436	0.436	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	58	42.03%	85.29%
							dis1	0		-
							dis2	3	2.17%	60%
							dis3	0		-
							dis4	0		-
							dis5	0		-
							dis6	3	2.17%	37.67%
							unidentified error or Not Answered	69	50%	44.01%
int(Ax^(n-1)/(Bx^n+G),x,a,b), a,b +ve; RNI	Integration\Rational functions\Logarithmic form	Explanation	126	126	0.5	0.5	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	63	50%	84.14%
							dis1	0		-
							dis2	4	3.17%	57.50%
							dis3	1	0.79%	40%
							dis4	0		-
							dis5	0		-
							dis6	1	0.79%	50%

							unidentified error or Not Answered	57	45.24%	49.88%
int(Ax^(n-1)/(Bx^n+G),x,a,b), a,b -ve; RNI	Integration\Rational functions\ Logarithmic form	Explanation	150	145	0.476	0.476	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	69	46%	85.90%
							dis1	0		-
							dis2	3	2%	60%
							dis3	1	0.67%	0%
							dis4	0		-
							dis5	0		-
							dis6	9	6%	52.22%
							unidentified error or Not Answered	63	42%	45.48%
int(ax^b,x); MC	Integration\Polynomials	Explanation	66	66	0.848	0.848	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	56	84.85%	67%
							No Constant	6	9.09%	30%
							Integration (xply b+1)	1	1.52%	0%
							Differentiated	2	3.03%	70%
							Integration (no div b+1)	0		-
							None Of These	1	1.52%	0%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,0,1); MC	Integration\Polynomials	Explanation	65	65	0.846	0.846	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	55	84.62%	69.02%
							Evaluation Of Limits	1	1.54%	0%
							multiplied by power plus one	1	1.54%	100%
							Differentiated	2	3.08%	20%
							Integration (xply b+1)	0		-
							None Of These	5	7.69%	24%
							Did Not Know	1	1.54%	0%
							Not Answered	0		-
int(ax^b,x,l,u)... 0 lt l lt u; MC	Integration\Polynomials	Explanation	78	78	0.859	0.859	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	67	85.90%	73.34%
							Integration (no div b+1)	2	2.56%	10%
							Integration (no div b+1) and sign error	0		-
							Differentiation	2	2.56%	30%
							Integration (div b)	0		-
							None Of These	7	8.97%	28.57%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,l,u)... fractional l & u; MC	Integration\Polynomials	Explanation	88	88	0.886	0.886	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	78	88.64%	65.22%
							Evaluation Of Limits	3	3.41%	26.67%
							Integration (no div b+1)	0		-
							Differentiated	2	2.27%	10%
							Integration (div b)	0		-
							None Of These	5	5.68%	40%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,l,u)... l lt 0 lt u; MC	Integration\Polynomials	Explanation	62	62	0.71	0.71	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	44	70.97%	74.43%
							Evaluation Of Limits	9	14.52%	15.22%
							Integration (no div b+1)	1	1.61%	40%
							Differentiated	3	4.84%	40%
							Integration (div b)	1	1.61%	60%
							None Of These	4	6.45%	55%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,l,u)... l lt u lt 0; MC	Integration\Polynomials	Explanation	70	70	0.771	0.771	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	54	77.14%	60.87%
							Evaluation Of Limits	9	12.86%	31.11%
							Integration (no div b+1)	1	1.43%	20%
							Differentiated	1	1.43%	0%
							Integration (div b)	1	1.43%	0%
							None Of These	4	5.71%	10%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,l,u)... u lt 0 lt l; MC	Integration\Polynomials	Explanation	87	87	0.816	0.816	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	71	81.61%	66.10%
							spurious factor of -2	1	1.15%	40%
							Integration (no div b+1)	2	2.30%	16.50%

							Differentiated	4	4.60%	30%
							Integration (div b)	1	1.15%	20%
							None Of These	8	9.20%	30%
							Did Not Know	0		-
							Not Answered	0		-
int(ax^b,x,-s,s); MC	Integration\Polynomials	Explanation	76	76	0.842	0.842	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	64	84.21%	66.83%
							Twice Evaluation 0-S	8	10.53%	15%
							Minus Twice Evaluation 0-S	1	1.32%	20%
							Evaluation 0-S Only	1	1.32%	0%
							Twice Derivative 0-S	0		-
							None Of These	1	1.32%	0%
							Did Not Know	1	1.32%	0%
							Not Answered	0		-
int(ax^b+cx^d,x); MC	Integration\Polynomials	Explanation	78	78	0.897	0.897	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	70	89.74%	63.53%
							No Constant	3	3.85%	26.67%
							Differentiated	3	3.85%	6.67%
							Integration (no div b+1)	1	1.28%	40%
							Integration (xply b+1)	0		-
							None Of These	1	1.28%	60%
							Did Not Know	0		-
							Not Answered	0		-
int(bx(exp(x/q)),x,b,q); MC	Integration\By Parts\Exponential functions	Explanation	77	77	0.455	0.455	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	35	45.45%	80.57%
							formula incorrect, sign.	13	16.88%	53.08%
							choose exp(x) to be v instead of dv	7	9.09%	51.43%
							integral of dv wrong, constant	12	15.58%	50%
							integral of dv wrong, no constant	2	2.60%	50%
							None Of These	7	9.09%	45.71%
							Did Not Know	1	1.30%	40%
							Not Answered	0		-
Int(exp(Bx)cos(Cx),x), B,C +ve; MC	Integration\By Parts\Exponential functions	Explanation	6	6	0.167	0.167	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	1	16.67%	80%
							Algebraic error	1	16.67%	60%
							Differentiation	2	33.33%	80%
							Integration	0		-
							error	0		-
							None Of These	1	16.67%	0%
							Did Not Know	1	16.67%	0%
							Not Answered	0		-
int(exp(px),x,a,b); MC	Integration\Exponentials	Explanation	185	185	0.703	0.703	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	130	70.27%	78.29%
							int(exp(nx))=exp(nx)	17	9.19%	36.47%
							int(exp(nx))=n*exp(nx)	18	9.73%	31.28%
							int like power	4	2.16%	38.25%
							guess	2	1.08%	60%
							None Of These	14	7.57%	44.29%
							Did Not Know	0		-
							Not Answered	0		-
int(exp(px/q),x,a,b); MC	Integration\Exponentials	Explanation	176	176	0.597	0.597	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	105	59.66%	76.42%
							int(exp(nx))=exp(nx)	15	8.52%	42.67%
							int(exp(nx))=n*exp(nx)	18	10.23%	45.56%
							int like power	5	2.84%	28%
							guess	2	1.14%	75%
							None Of These	31	17.61%	55.26%
							Did Not Know	0		-
							Not Answered	0		-
int(exp(x/q),x,a,b); MC	Integration\Exponentials	Explanation	182	182	0.582	0.582	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	106	58.24%	80.47%
							int(exp(nx))=exp(nx)	13	7.14%	39%
							int(exp(nx))=n*exp(nx)	33	18.13%	45.55%
							int like power	2	1.10%	0%
							guess	0		-
							None Of These	27	14.84%	40.70%
							Did Not Know	1	0.55%	0%
							Not Answered	0		-

Int(x^n+exp(-mx),x,A,B), A, B +ve; NI	Integration\Exponentials	Explanation	172	172	0.459	0.459	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	79	45.93%	77.82%
							Wrong	93	54.07%	40.75%
							Not Answered	0		-
Introduction to tests; WI	authoring templates \Introduction screens	Explanation	769	769	0	0	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							yes	760	98.83%	60.31%
							no	8	1.04%	6.25%
							Not Answered	1	0.13%	0%
x*exp(bx); MC	Integration\By Parts\Exponential functions	Explanation	81	81	0.679	0.679	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	55	67.90%	83.64%
							integrated each term separately	3	3.70%	53.33%
							differentiated exp(Ax) instead of integrating	8	9.88%	27.50%
							incorrect sign in the formula	8	9.88%	42.50%
							differentiation instead of integration	0		-
							None Of These	6	7.41%	56.67%
							Did Not Know	1	1.23%	40%
							Not Answered	0		-
x^2exp(px); MC	Integration\By Parts\Exponential functions	Explanation	72	71	0.38	0.38	**Outcome name**	**Times answered**	**Percentage of times answered**	**Mean for outcome**
							Correct	27	37.50%	78.52%
							integrate term by term)=exp(nx)	8	11.11%	42.50%
							int(exp(nx))=n*exp(nx)	8	11.11%	40%
							int like power	9	12.50%	60%
							guess	1	1.39%	20%
							None Of These	17	23.61%	52.94%
							Did Not Know	1	1.39%	60%
							Not Answered	1	1.39%	60%