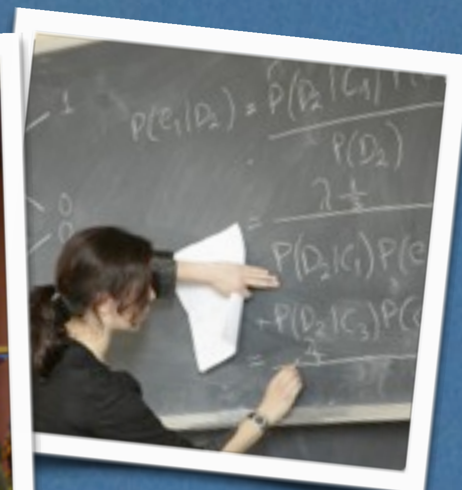
MAPPING UNIVERSITY MATHEMATICS ASSESSMENT PRACTICES



Edited by
Paola Iannone
Adrian Simpson

Mapping University Mathematics Assessment Practices

Edited by

Paola Iannone
University of East Anglia

Adrian Simpson
Durham University

Mapping University Mathematics Assessment Practices
Published 2012.
University of East Anglia
ISBN 978-1-870284-01-1

Foreword

In 2011, a summit was held to discuss various aspects of the state and future of the mathematics curriculum in higher education in England and Wales. Among the topics for debate were students' perceptions of their degrees, essential skills of mathematics graduates and the value of subject knowledge, memory and technical fluency for students on mathematics degrees (Rowlett, 2011). One particular theme was how we assess what we claim our degrees deliver, particularly at a time of major changes to higher education in England and Wales (Levesley, 2011). Questions of efficiency of assessment, its validity, its perception by students and the mix of methods were discussed. The summit report also echoed the concerns expressed by others (e.g. LMS, 2010) about the ownership of choices made about assessment patterns.

The *Mapping University Mathematics Assessment Practices* Project (MU-MAP) was developed to examine the current state of assessment in our undergraduate degrees. It was designed not only to give a broad overview of practice, by looking across our higher education institutions, but also to have an eye to the future and alternatives. Our focus throughout is on summative assessment: while formative assessment is of great importance, the project was intended to explore the methods we use to make public judgements and statements about performance and attainment.

It should be noted that we have chosen to avoid the word 'innovative' with respect to alternative assessment methods. We have done this for a number of reasons. The word implies a freshness which may not be the case: some of the alternatives we identify have been common practice in some institutions for decades, even if they might be new to others. The word also implies an improvement and one of the issues which emerges from the survey reported in the first chapter is that not everyone agrees that alternative methods are better. Moreover, as one head of department noted, there is a discourse related to the use of 'innovative' which suggests a lack of reflection, a conservatism and some inertia in mathematics departments, and our survey failed to find evidence of these.

As chapter 1 notes, the closed book examination is by far the most dominant assessment method across the sector and discussions with heads of department suggest that this is not simply the result of conservatism, but a belief that written examinations strike an appropriate balance between efficiency, validity and fairness – they are a 'gold standard' against which other assessment methods must be judged. However, it is also clear that a number of alternative methods are in quite widespread use: projects, presentations, coursework, online quizzes, etc. Part II of this book outlines an opportunistically collected sample of these methods obtained by asking departmental heads and directors of teaching about who was using alternative methods in their departments. The reasons for developing these alternatives vary from reducing workload, through testing different types of skills to addressing institutional requirements.

Part III of the book delves deeper into the issues surrounding the implementation of alternative assessments. By funding a small number of evaluation studies, the

MU-MAP Project allowed colleagues to address the extent to which alternatives did indeed achieve their intended outcomes. These studies explore whether approaches as varied as multiple-choice tests for assessing students' understanding of proof, peer judgement of pairs of scripts or oral examinations can address the concerns raised in the 2011 HE Mathematics Curriculum Summit.

References

Levesley, J. (2011) Taking control of the assessment agenda. In Rowlett, P. (Ed) *Report of the HE Mathematics Curriculum Summit.* York: Higher Education Academy, 21-23.

London Mathematical Society. (LMS) (2010) Mathematics degrees, their teaching and assessment. Accessed via `http://futurelms.wordpress.com/2010/04/08/teaching-position-statement/`

Rowlett, P. (2011) *Report of the HE Mathematics Curriculum Summit.* York: Higher Education Academy.

Contents

Part I
Patterns of Current Assessment Practice

The first part of this book gives an overview of the current state of assessment in Higher Education institutions in England and Wales. The first chapter summarises the results of an extensive survey of assessment methods in mathematics departments. Using publicly available data or information kindly supplied by departments, the chapter outlines the key assessment methods used and draws out connections between methods and topics. It demonstrates clearly that the closed book examination remains by far the most prevalent method of summative assessment. Discussions with staff suggest that this is not entirely a matter of conservatism, but a genuine belief in the written examination as having the best available combination of qualities.

The second chapter reports the findings of a literature review on assessment in mathematics at university level. It demonstrates that, while there is clear activity in this area, the majority of the work consists of reports of alternative practice and there is very little empirical work which seeks to compare assessment methods or carefully research the impact of assessment methods on approaches to learning, efficiency, validity etc.

Chapter 1
A Survey of Current Assessment Practices

Paola Iannone and Adrian Simpson

1.1 Introduction

In this chapter we present the results of a broad survey of assessment methods currently used in mathematics departments in England and Wales. We draw on publicly available data, as well as information generously supplied by many of the departments about their modules and assessment methods. The chapter aims to give an indication of the general trends in the use of different forms of assessment, an overview of the connection between the content and aims of modules with the assessments used and, through an analysis of interviews with heads of departments and directors of teaching, some of the rationale for the current patterns of assessment.

1.2 Methods

To get a sense of the current state of assessment in undergraduate mathematics, we developed a systematic approach to gathering data from higher education institutions. The on-going changes to the higher education system mean that universities are being required to be more open about a whole range of "Key Information Sets" – data on teaching activities, fees, student satisfaction, etc. In particular, universities are being required to make public information about their assessment methods. Even though these requirements were not yet fully in force at the time of the survey, most universities have already provided a large amount of data online about their degree programmes.

We identified, first, a list of mathematics departments in England and Wales (taken from a publicly available league table) and a list of the degree courses they

Paola Iannone, e-mail: p.iannone@uea.ac.uk
School of Education and Lifelong Learning
University of East Anglia

Adrian Simpson e-mail: adrian.simpson@durham.ac.uk
School of Education
Durham University

offered. In many cases, the information we needed was readily available and in others, departments generously provided the information. In only 6 cases did our enquiries result in insufficient information to include the department in our analysis at all and for a further 8 there was not enough data to undertake some of the quantitative analyses or we felt comparisons were more difficult to justify. Given that many universities offer a range of different degrees with a significant mathematics component, for each institution we identified the degree course which most closely resembled a stereotypical three-year BSc in mathematics.

Within this, we identified the modules taught within the mathematics department (including optional modules, but discounting modules provided by other departments). We recorded the module title, year in which it was normally taken, its contribution to the mark for the year (normally in terms of the ratio of the credit value for the module to the normal number of credits needed for the year), its contribution to the final degree classification, the assessment methods and the contribution of each assessment method to the final mark for the module.

It is commonplace to make a clear distinction between formative assessment (which emphasises students' and lecturers' understanding of current performance to support potential changes to learning and teaching activities) and summative assessment (which is intended to reflect the outcome of the learning that has taken place). Most departments include some assessment which is entirely formative in nature, but as noted below, there are many assessment activities undertaken which both provide students and lecturers with on-going information and contribute to a final mark for a module. In this survey, the focus has just been on summative assessment, so the data contains information only about assessment which contributes to a module mark. We included situations where the module or year mark does not directly contribute to a final degree classification - in fact, in the sample, only seven institutions explicitly state that the first year mark contributes to the final degree and in the vast majority of universities the first year is 'qualification only'. In the end, sufficiently robust data for quantitative analysis was obtained for 43 degree courses, involving 1843 modules.

In addition to this data, we contacted departments to ask if a senior member of staff (the head of department or director of teaching) would be able to take part in a telephone interview about assessment on their degree programmes. This resulted in 27 interviews, each around half an hour in length, which focussed on trends in assessment, different types of assessment practice, rationales for patterns of assessment and the interviewee's personal views on these assessment practices.

1.3 Analysis and findings

Investigating the 1843 modules, it became apparent that the closed book examination is the most dominant assessment method. Over one quarter of the modules (535) in the sample are assessed entirely by closed book examination and nearly 70% of the modules (1267) use closed book examinations for at least three quarters of the final mark.

Looking across institutions, we see that the dominance of closed book examinations varies according to year group. There is a general trend in which closed book examinations play a larger role in later years. Figure 1.1 shows the proportion of closed book examinations (averaged simply across all modules at all universities) for each year. In addition, because most universities include a final year project which has no examination component, the figure includes a measure of the average proportion of closed book examinations within final year modules, having removed any module explicitly described as a final year project.

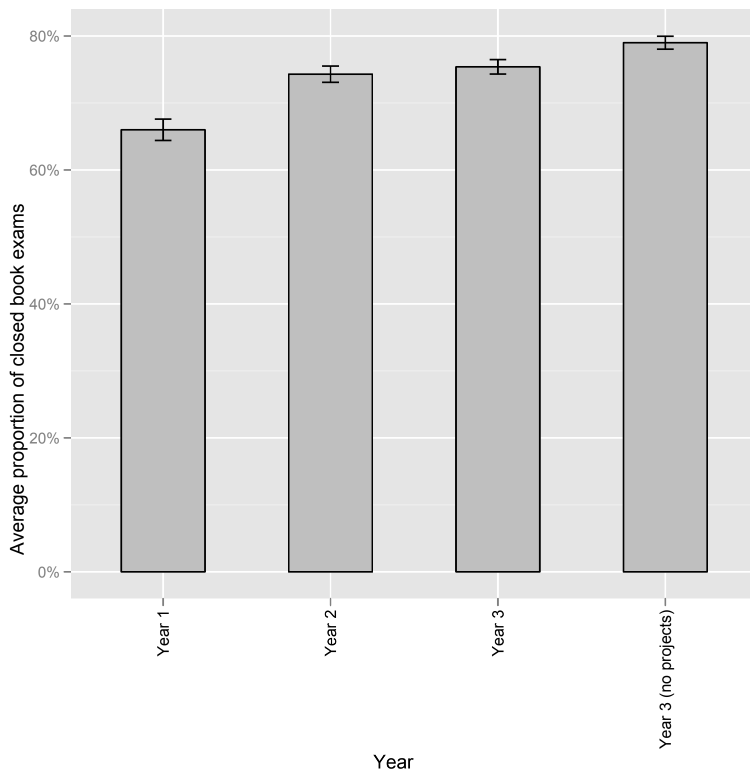


Fig. 1.1 Average proportion of closed book examinations in modules by year

While this gives a view of the situation across the country, to gain a sense of how much the closed book examination can dominate the final degree result in individual institutions we used a method outlined in Iannone and Simpson (2011). For each institution, we averaged across all the modules on offer, weighted for the contribution the module makes towards the degree mark. Figure 1.2 gives the result for each institution, against their ranking according to a publicly available league table of mathematics departments.

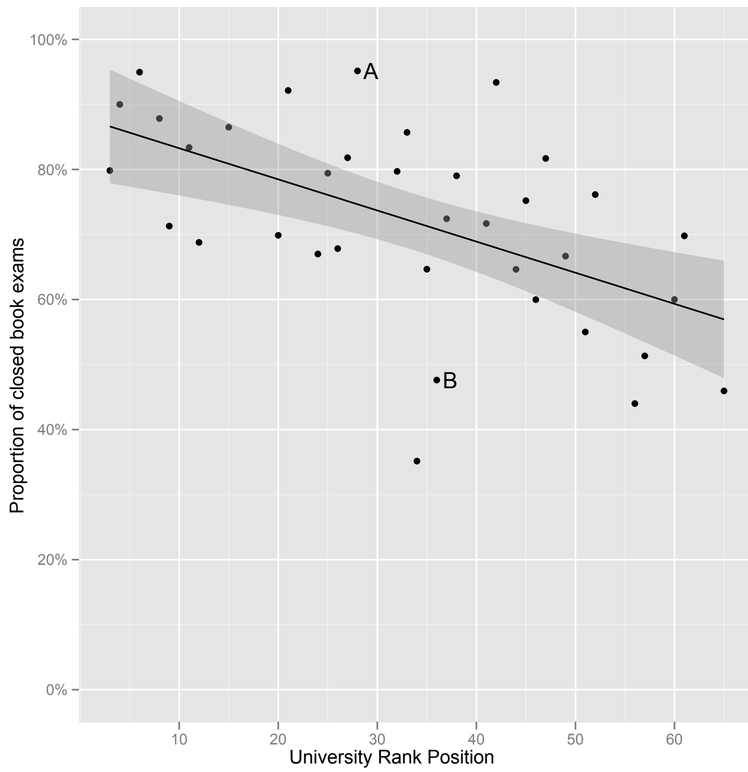


Fig. 1.2 University rank position and proportion within the final degree classification mark assessed by closed book examination

A number of things stand out from the graph in Figure 1.2. First, the dominance of the closed book examination is clear; in only four institutions is the average weighted contribution of examinations less than 50% and the median contribution is 72%. Second, the “line of best fit“ (with 95% confidence interval) gives a general sense of a relationship between the standing in the league table and the use of examinations. There is a significant inverse relationship ($\rho(39) = -0.571$, $p < 0.001$) between the weighted average proportion of closed book examinations and the institution’s league position. Third, there are some notable outliers: institutions which, for their league table position, significantly over or under use closed book examinations. It should be noted that the method for calculating the proportion of closed book examinations can lead to some exaggerated figures: a university which has, for example, a very large number of options all assessed exclusively by examinations, but has a compulsory project worth a substantial amount towards the final mark may have the influence of that project downplayed when the average across all the modules is taken. However, as we shall see below, drilling down into the data at particular points does suggest there are genuine differences behind the picture.

Indeed, to gain a sense of the variation in the pattern of assessments across these institutions, we adopt a case study approach by choosing two outliers which have similar positions in the ranking, are both pre-92 institutions and have similar entry requirements (AAA for University A and AAB for University B). We examine the different approaches to assessment on their undergraduate degree courses.

1.4 Case studies

We will call the two outliers Universities A and B (marked in Fig 1.2). They were chosen because, as well as being outliers and having some broad similarities as institutions, both of these departments provided a complete set of data for their mathematics modules and both heads of department kindly gave their time for interviews.

1.4.1 University A

University A offers both 3-year (BSc) and 4-year (MMath) degree courses in mathematics. Marks accrued in year one do not count towards the final degree classification in this university. As part of its BSc in mathematics it offers 44 modules ranging over applied and pure mathematics, statistics and probability and financial mathematics. In the first year, students take five modules, with no options available. Three of these are assessed by 3 hour final examination alone. One has a multiple choice test worth 30% and a final exam, and the other has a group project worth 10% and a final exam.

In the second year, students take two compulsory modules, both assessed solely by closed book examinations, and choose from eleven optional modules (with an opportunity, if they wish, to take one module from outside the department). All of these are assessed solely with an examination.

In the final year, with the exception of a compulsory project, students can choose from 36 different modules (again with an option of one module from outside the department). These are generally either 100% examination based or have a small coursework component. It is clear that, the final project aside, the department places great emphasis on the closed book examination.

In the interview with the head of department, it appeared that there had been a reasoned move away from coursework. There is a wider variety of formative assignments, but concerns about plagiarism and collusion mean that weekly coursework sheets are increasingly being replaced by in-class tests or no longer carry credit towards the module mark. There was no feeling that the department was particularly conservative: one member of staff was experimenting with the use of Maple and quizzes delivered online and there has been some discussion about introducing oral examinations or other forms of assessment. However, the concerns about plagiarism appear to have been the main driver in moving the department to its current position.

1.4.2 University B

University B also offers 3-year and 4-year degree courses in mathematics. In common with most universities, the first year of the degree course does not accrue marks toward the final degree classification. The pattern of the curriculum in the first year is very similar to University A. There are six compulsory modules (four of which are half the credit rating of the other two) and the opportunity to take two modules of options (including one from outside the department).

The difference comes in the methods of assessment: while four of the compulsory modules are 80% examined, with 20% coursework, normally in the form of weekly homework sheets, one has 40% coming from weekly homework sheets and Maple assessments and another has no examination at all (with assessment coming from weekly sheets, tests and a portfolio of computer practicals). The options within the mathematics department vary from 100% project based modules to those assessed by 50% examination and 50% coursework.

In the second year, students have to take 70 (out of 120) credits of compulsory modules with the remaining 50 from options given in the mathematics department. No module has more than 80% of its assessment based on examinations, with the modal pattern of assessment including 20% from weekly homework sheets. Two modules (one option and one core) are 100% assessed by a project and presentation of that project. Other modules have half of their credit coming from weekly sheets and computer tests.

The university is unusual in not having a module which is clearly labelled as a compulsory project in the final year. Indeed, there are no compulsory courses, but a choice of 20 modules within the mathematics department (or 12 in a related department). Only five modules are 100% examined (mostly in pure mathematics), others vary with the modal pattern of assessment being 80% examination and 20% coursework. However, there are a number of modules with no examination component, including research skills modules, history of mathematics and a choice of (non-compulsory) projects in various areas.

The head of department was clear that the move away from examinations, while gradual and on-going, was a positive decision on the part of the department. There is a concern amongst the staff that examinations have become too much about reproduction and the move away from them allowed the department to focus on the development and assessment of a wider range of skills. Moreover, they perceive a pressure to find their position in a market amongst universities and mathematics departments and to ensure that their students have a competitive advantage with employers. Thus the emphasis has moved away from education based predominately on developing the next generation of mathematicians. This had led to more project work, more use of packages such as Maple and may in future lead to other changes such as more presentations and the use of oral examinations. The head of department described a department that is content with the changes, though some are more comfortable with experimenting with assessment methods than others.

1.5 Alternatives to the closed book examination

Having seen the dominant role played by closed book examinations - a role which increases as the value of modules to the final degree classification increases - it is also worth exploring what the main alternatives are and what areas within the mathematics curriculum tend to use those alternatives.

Exploring our data, a number of key areas in which closed book examinations played a lesser role emerged. These included final year projects, statistics and financial mathematics modules, computing, problem solving, history of mathematics, mathematics education and employability skills.

1.5.1 Projects

Of the 42 universities for which we have robust data, 32 have a final year project and many of these are compulsory. Many of these are the equivalent of a double module for the students and they are expected to work on them across the whole of the final year. Thus, for what is often a rare piece of non-examination based assessment, these projects have a considerable influence on the final degree mark. More than half of these projects (20) are assessed purely on the submission of a written report, but many others include credit for an interim report, a presentation (either as a poster or a verbal presentation) or an oral examination component. In general though, the written component is the majority portion (the average across all the projects being 75%) with presentations the next largest component (where they are used, they average 17%).

1.5.2 Statistics and financial mathematics

In our sample, 301 modules appeared to be in an area of statistics. While still generally dominated by examination (first year modules and final year modules have a similar proportion of their marks coming from examinations, but second year modules have a smaller proportion), there is evidence of a slightly different pattern of assessment. There is more investigatory work, use of computer packages for work on large data sets and written reports. In addition, a number of the examinations are “open book” (allowing anything from statistical tables to textbooks and other materials to be taken in to the examinations).

Similarly, for the 61 financial and business mathematics modules, there is evidence of more use of reports and presentations; one such module even uses in-class electronic voting systems as part of the summative assessment. However, many modules also have a heavy reliance on closed book examinations.

1.5.3 Computing

There are 51 modules in computing or programming across the sample. Some of these are traditional courses teaching a programming language (such as Fortran) but many are modules using mathematics packages like Mathematica or Maple. Understandably, these rely on a far smaller proportion of closed book examinations, with practical programming tasks as the main alternative.

1.5.4 Problem solving

There is a relatively small number of modules (11) with titles such as “problem solving” or “investigations”. The emphasis of these modules appears to be on developing mathematical problem solving skills rather than on the learning of a specific area of mathematical content. These tend to be assessed either by a single report of a substantial investigation (in which case, they are more akin to a project) or by a number of smaller investigations submitted as written reports of students’ findings as they work on each problem.

1.5.5 History of mathematics

There are 12 modules in the history of mathematics across the sample, all but one in the final year. There is quite a wide variation in the assessment: five of them have the majority of the marks coming from final examinations, the others are dominated by essays (in the form of either a single, high stakes essay or a number of essays submitted throughout the module).

1.5.6 Mathematics education

There are 25 modules in mathematics education, 20 of these in the final year. Only four of these have a closed book examination component (and here these tend to be minority components). These modules tend to include many different assessment components, including reports on school visits, essays, projects, presentations, etc. One such course has six assessment components: a project, a presentation, classroom log, end of module report, an evaluation of classroom performance and a key skills assessment.

In none of these areas should it be all that surprising that there is less use of closed book examinations and there are obvious reasons for the approaches taken. For example, in statistics the need to assess the practical skills of analysing large data sets suggests the use of projects and computer packages and the dual concerns with both application and derivation of formulae suggest a mix of open book and closed book examinations. Similarly, where the work is highly individualised - such as with projects, problem solving or mathematics education (containing visits to different schools by different student) - closed book examinations are less appropriate than essays and projects.

Of course, the list is not exhaustive. There were a number of less common topics with different assessment patterns. For example, a small number of institutions have explicit study skills, personal development or employment skills modules which tend to be assessed with coursework, essays and projects and some have mathematical modelling courses assessed with group projects, posters and presentations.

It is also not as straightforward as having different assessment methods tied to particular types of module content. As Part II of this book notes, there are alternative assessment practices in many different areas and many individual modules stood out in our sample as having assessment forms quite distinct from others of the same name. These include for example, a linear algebra course with 40% of the mark coming from computer practicals with Maple; a non-linear systems course assessed with two pieces of coursework and a project and a differential equations course that contains individualised projects.

1.6 Interviews with heads of department

As part of the survey, for each department we attempted to contact either the head of department or a person with a similar overview of the assessment patterns in the undergraduate degree, such as the director of teaching. In total 27 people were interviewed over the telephone, each interview lasted on average half an hour and field notes were taken. The interviews were structured around a number of themes: what trends were seen in patterns of assessment, what traditional and alternative methods of assessment were employed, what lecturers felt about the assessments used, and what areas would the individual wish to change. In addition interviewees were asked to nominate anyone using any notable alternative assessment practices (many of which formed the basis for the case studies in Part II of this book).

The field notes were examined for issues, particularly for those which were discussed by a large number of interviewees. These issues were drawn together into six key themes which emerged from the data as we sought patterns across the issues (Bryman and Burgess, 1994). The main themes which emerged were about conflicting trends in assessment practice; fairness, plagiarism and collusion; contentment with existing patterns; reasons for change (which included efficiency and institutional pressures); employability and driving student learning.

Across the interviews as a whole, there was an interesting sense of conflicting trends. In most cases, the emphasis was on a movement to decrease the proportion of formal, closed book examinations (though often in a context in which they would remain dominant), but in others there was a movement away from coursework, essays and projects. This latter trend was justified by concerns about fairness, plagiarism and collusion.

The issue of fairness was raised in relation to assessment of essays and projects. These are sometimes marked across a number of different assessors and may have mark schemes which are less detailed than those used for other work in mathematics. Hence some people raised the question of how to make sure that similar marks are being awarded for work of similar quality. The issue of plagiarism and collusion was discussed in many interviews both in relation to projects and essays and in relation to coursework sheets. Often people explicitly raised concerns about weekly homework sheets and how they could be sure that the work represented the students' own efforts. It was also noted that these types of assessment made it very hard to identify plagiarism (since it would be quite natural for students to use similar methods and even similar variable names in their solutions to the same problem). In the case of essays and projects it was noted that some of the issues can be addressed by plagiarism detection software.

The issue of students working together was raised, in some cases as an entirely legitimate way of working and in others as potentially undermining the concept that marks should reflect the work and understanding of an individual. As some interviewees noted, the attribution of small amounts of summative credit to an assignment can increase engagement with it, but the corollary of this was also noted: a constant emphasis on marks can lead students to lose sight of the importance of learning the material for themselves. There is an interesting tension to be noted here regarding assessment which is overtly described as group work: some described this as an important employability skill which a degree should provide, while others worried about how to measure a student's individual contribution.

Where these issues have been a concern, some departments are moving away from the credit bearing weekly homework sheet towards in-class tests or reconfiguring them as formative assessments, as, for example, in University A discussed above.

Many of the people interviewed noted that there was general agreement across the staff in their department with the existing pattern of assessments. Only one head of department explicitly noted tension between those who wanted to move the balance away from closed book examinations and those who wanted to maintain it (or increase it). That said, where alternative assessment methods were in place, there was uniform support for them. In most cases, the development of these alternatives came from individuals wishing to try something out in their own modules and there was every sense that they were supported across the whole department.

The reasons for introducing alternatives seemed to vary. In some instances it was efficiency. Many people noted the large workload associated with assessment, with the general comment that projects, essays, weekly homework etc. tend to be more time consuming than examinations to mark, but even examinations were considered a significant burden. This was particularly the case where it was noted that student numbers have increased. In some cases, the intake has more than doubled in three years, making previous methods of assessment unsustainable. This issue was raised most commonly in relation to introducing computer based assessment where packages such as Maple TA or systems designed in-house can be used to assess and provide feedback to large numbers of students with a much reduced workload for lecturers.

In a relatively small number of interviews, one of the reasons given for change was institutional pressure. Universities were seen as influencing departmental strategies by pushing for more coursework or more explicit assessment of employability skills. External examiners were also seen as a source of pressure: in some cases pushing for change or in others advising against an alternative method such as an oral examination. There were a couple of instances of student pressure: in one case the introduction of peer assessment was abandoned after students raised concerns about whether peers were qualified to make robust judgements; in another, a perception that students were bored with a restricted assessment diet had led to the introduction of some alternative methods.

The issue of employability skills was another theme which emerged. Some justified alternative forms of assessment as more realistic of the kinds of tasks students would encounter in later employment and others noted the need to be able to market their degree to students as giving them an employment advantage over others.

Remarkably few people discussed the students' perceptions of assessment and the extent to which assessment might drive learning (with the exception, noted above, that the thirst for grades may lead to plagiarism or collusion with coursework). Those that did linked the choice of assessment methods with the need to have the best measure of an individual's understanding of the material. In most cases, this was tied to the closed book examination, but occasionally to the final year project or to the idea of an oral examination.

1.7 Discussion

The survey gives a sense of the current state of assessment across the country – albeit a snapshot of a system in a state of flux. Not least because of the on-going changes to higher education funding, the expectation is that assessment will continue to be a substantial matter of concern for all.

The image developed by the survey is of a system dominated by the closed book examination, but – from the point of view of heads of department – not inappropriately so. The closed book examination is seen to have an appropriate balance of validity, efficiency, reliability and fairness. The differences between departments are differences of degree rather than substance: the median contribution of closed book examinations in modules counting towards the final degree was 72% and few departments have closed book examinations accounting for less than 50% (when averaged across all their modules).

What differences we did discover between departments may be an artefact of the range of topics they teach. As noted in section 5, alternative assessment methods are much more appropriate for topics such as statistics or the history of mathematics. However, there may also be an influence on the range of assessments used which comes from the aims of the particular department: those who see their emphasis skewed towards students aiming for further study or producing the next generation of mathematicians may choose a different pattern from those whose emphasis is skewed towards students wishing to use their degrees for employment outside academia.

The pressures for change appear to come predominately from two areas: a concern for efficiency in an expanding sector and, to a lesser extent, institutional pressures. However, the drivers of change appear to be dominated by the committed individual. While heads of department noted widespread support and, in some cases, changes being developed after department wide discussion, most alternative methods appear to have been developed by individuals interested in making a change in their own modules. This suggests that the concerns of the LMS (2010) about the control of teaching and assessment and those of Levesley (2011) about an inherent conservatism are not wholly supported by the data we have collected. That is not to say that, as the higher education system changes dramatically in the next few years, we should not be wary of these issues escalating; but at the moment the locus of control seems grounded in departments and the changes come about from committed and interested individuals working in generally supportive environments.

Acknowledgements

We would like to thank Oana Radu for her role in collecting much of the data for this chapter.

References

Bryman, A., & Burgess, R. G. (Eds.). (1994) *Analyzing qualitative data*. London: Routledge.

Iannone, P. & Simpson, A. (2011) The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*. 30(4), 186-196.

Levesley, J. (2011) Taking control of the assessment agenda. In Rowlett, P. (Ed) *Report of the HE Mathematics Curriculum Summit. York: Higher Education Academy*, 21-23.

London Mathematical Society (LMS). (2010) *Mathematics Degrees, their Teaching and Assessment*. Accessed via `http://futurelms.wordpress.com/2010/04/08/teaching-position-statement/`

Chapter 2
A Review of the Literature in Undergraduate Mathematics Assessment

Oana Radu

2.1 Introduction

In this chapter we present the results of a literature review of mathematics assessment at university level. We discuss the methods of the review, the classification we adopted for the papers we included and we synthesise the outcomes of the review.

2.2 Methods

We conducted an online search that included the following databases: Education Resources Information Center, JSTOR Arts and Sciences Collection I-IV, EBSCO EJS, British Education Index, Oxford Reference Online and Google Scholar. We also searched within professional journals such as *MSOR Connections* and *Teaching Mathematics and its Applications*. The search was conducted between November 2011 and February 2012. We used the following key words: 'mathematics', 'assessment' and 'university'. Results of the search were then screened for relevance, and we excluded any paper which was not directly concerned with assessment of mathematics at university level.

The 78 papers we found from this search were subsequently grouped into three categories: theoretical (18), empirical (5) and professional papers (55). We adopted the following definitions for this subdivision: *Theoretical papers* are papers which discuss assessment of mathematics at university level from a theoretical viewpoint, often drawing from results found in the general assessment literature in higher education and which do not report directly on new empirical work. *Empirical papers* report results of empirical studies on assessment at university level. *Professional papers* are usually shorter articles written by practitioners (e.g. lecturers of mathematics or statistics) describing the implementation and outcomes of some assess-

Oana Radu, e-mail: `o.radu@uea.ac.uk`
School of Education and Lifelong Learning
University of East Anglia

ment method, often in their own teaching, teaching reflections or descriptions of the design and use of novel technological tools.

2.3 Theoretical papers

2.3.1 Factors affecting change in assessment practices

There are a number of issues raised in the literature which appear to be a root cause of change. One such issue is change to resourcing.

Cuts to universities' budgets can affect the delivery and structure of modules. Such financial constraints put pressure on universities and departments. Moreover some mathematics departments have lost their function as providers of service modules, as departments such as engineering and business consider that offering mathematics modules from their own staffing will benefit students more than having them trained by mathematics departments (Smith and Wood, 2000; Houston, 2001).

In a higher education system seen increasingly as market driven, universities often reorganise their courses in an effort to make them more appealing and recruit more students. This strategy contributes to restructured assessment practices. It is also noted that some forms of assessment are more difficult to implement with wider access to higher education and the increase of class size. Thus, technology has the potential to make assessment practices more effective and reduce lecturers' workloads; it can cater for a wide audience and makes mathematics readily available to students with special educational needs or for distance-learning.

Assessment is also tailored according to the market force demands for graduates. Mathematics graduates find employment in different fields. As such, universities tend to adjust assessment strategies and aim to enhance students' employability skills such as critical thinking (Anderson et al., 1999; Smith and Wood, 2000; Houston, 2001). It is argued that examinations do not reflect either the forms of argument used by mathematicians in their professional work or more general workplace practices (Houston, 2001).

A varied assessment diet (e.g. comprising open-ended problems, essays, portfolios, computer simulations, multiple-choice questions, computer-based assessments, problem sheets, presentations) addresses employability skills and educates students to become efficient players in the workforce (Linn, Baker and Dunbar, 1991).

More realistic mathematical practice has ample opportunity for discussion, reflection and interaction. Thus, assessment strategies should reflect the way in which mathematicians work but should also equip students with the necessary skills required within the workplace (Houston, 2001). Assessment should not be used solely to inform lecturers about students' knowledge and understanding, but should also develop students' skills and abilities to self-reflect (Houston, 2001).

Assessment in mathematics can appear to some as a conservative practice. Some lecturers feel that traditional assessment methods such as written exams and weekly coursework best assess students' mathematical knowledge. Students' opposition to new practices is also seen as a barrier to change, as sometimes they feel it is more convenient to engage in familiar mathematical tasks than engage with new practices. Among the apparent obstacles to implementing new assessment practices we found reference to lecturers' ignorance of novel forms of assessment or lack of resources (e.g. time, money, energy) (Burton and Haines, 1997; Huston, 2001).

2.3.2 Benefits and shortcomings of using alternative assessment strategies

Computer Assisted Assessment (CAA) can facilitate various tasks, such as access to large problem databases and test prototypes, instant marking facilities, immediate feedback or access to statistical packages to analyse students' performances (Engelbrecht and Harding, 2005a, b). Using computer-based assessment also reduces marking and administration time. This technology increases lecturers' opportunities to tap into and use large problem databases. Moreover, lecturers have the opportunity to be directly involved in designing problems (Sangwin, 2003).

Using technology for evaluating mathematical tasks provides a series of benefits. In the literature, several CAA systems are discussed. For instance, AIM (Alice Interactive Mathematics) does not penalise incorrect mathematical language errors (Sangwin, 2003). AIM offers students the possibility to validate answers before being marked. Furthermore, students have the opportunity to revisit, review and practise as much as necessary. Mathwise is a multimedia system used for mathematics assessment. It offers the possibility to assess calculation processes within a problem and not just the final answer (Pitcher, Goldfinch and Beevers, 2002). The evidence suggests that students feel they work equally well within environments such as Mathwise and traditional test situations. They appear to enjoy the computer environments as these provide the chance to practise as much as needed and to receive instant feedback (Pitcher, Goldfinch and Beevers, 2002).

Technology is not yet as advanced as to automate all mathematical tasks (Engelbrecht and Harding, 2005a, b). For instance, it is not yet possible to assess online proof writing or reasoning skills. In some instances, it is also difficult to offer part marks for students' solutions to mathematics problems (Sangwin, 2003). As an example, Pitcher et al. (2002) argue that Mathwise can be improved by creating more detailed feedback when students give incorrect answers.

Fairness, transferability and generalisability, content quality, complexity and coverage of assessment are also discussed in these papers. When introducing new assessments strategies, it is essential to consider the relevance of the criteria used and to match the criteria of assessment to the learning outcomes and focus of the module (Linn, Baker and Dunbar, 1991). While assessing group work (e.g., projects or presentations) it is essential to differentiate the contribution of each team member in the fairest way possible.

2.4 Empirical papers

Empirical studies are designed to provide an evidential base for making decisions about assessment methods. They can be driven by the desire to develop differentiated performance assessments (Stull et al., 2008), to investigate the relationship between student performance, past mathematics experience and perceptions of statistics education (Cybinski and Selvanathan, 2005), to integrate technology into the teaching and learning of statistics (Cybinski and Selvanathan, 2005) or to introduce concept maps in teaching mathematics (Mwakapenda, 2003).

Explorations of traditional and alternative forms of assessment in mathematics and statistics (Iannone and Simpson, 2011) found closed book examinations dominate the assessment diet in undergraduate mathematics. Alternatives to closed book examinations tend to consist of combinations of small projects and closed book examinations, open book exams, presentations (either individual or in groups) or written work (e.g. reports, logbooks, essays). Combinations of computer-based assessments, projects, multiple-choice questions, problem sheets, reports or presentations may also be used to assess modules.

Mwakapenda's (2003) study explored first year students' understanding of algebraic, numerical, graphical and geometrical domains through the use of concept maps. Interviews exploring students' understanding were used to investigate the types of mathematical examples used in creating links between mathematical topics. Research findings indicate that students struggle to express their understanding of conceptual links between mathematical concepts. Mathematical concepts seem to be understood in the context of a taught module only. Students seem unable to connect notions acquired in different mathematics modules.

Stull et al. (2008) investigated a type of formative assessment with four groups of students on a module in differential equations during two semesters. The formative assessment consisted of increased feedback on the quizzes taken by the students (who were not told about the study). Each group contained about 30 students. Similar quizzes and materials were given to each group. The lecturer chose to assign either a mark or detailed feedback on groups of papers. The study consisted of three groups that received formative assessment and one control group. 79 students participated in the study. Overall, 58.2% of the students performed higher than expected. Subsequently, the data was divided into three groups: high achievers (80th percentile and above), students who achieved well below the expectations (20th percentile and below), and mid range achievers. Being exposed to a number of quizzes during a semester influenced students' performances. Students were also asked to estimate the weekly number of hours spent engaged in academic activities, going to class, doing homework as well as the number of hours spent interacting with friends. Researchers noticed that at the end of the semester the number of hours spent on homework increased while the time spent interacting with friends decreased. At the beginning of the study, the students who performed below predictions agreed that they spent twice as much time working in a part-time job than their colleagues who scored higher. Towards the end of the semester, these students increased the hours spent doing homework and decreased the hours dedicated to their part-time jobs.

This may be viewed as 'catching up' in the last part of the course. The paper concludes that it is essential to have a good start to the semester and a work plan.

Cybinski and Selvanathan (2005) investigated the benefits of using flexible learning. This strategy offers students greater responsibility for and flexibility in their own learning process. One group of students enrolled in an introductory statistics module was exposed to the traditional lecturing style, consisting of weekly two-hour lectures, a one-hour computer based session and one tutorial hour. The second group was exposed to a flexible learning environment where students had access to a web-based learning tool, a one-hour computer session and optional problem-based workshops. The results of the survey investigating the influence of each learning mode on students' performances suggest that the group studying under a flexible learning environment experienced a high level of test and performance anxiety. Thus, a flexible learning environment combined with minimal direct teaching may not be the most beneficial learning environment for students.

2.5 Professional papers

The majority of mathematics professional papers discuss technology tools such as GeoGebra, the Computer Algebra Based Learning and Evaluation (CABLE), XML, MathML, Mathletics (a suite of computer-assisted assessments), or the System for Teaching and Assessment using a Computer Algebra Kernel (STACK). The remaining papers describe mathematics assessment methods other than closed book examinations.

Lecturers are driven by the desire to offer students alternative strategies for learning mathematics or by the desire to increase students' abilities to formulate and conceptualise mathematical ideas. Assessment tools incorporating technology-based materials help students write and explicitly formulate mathematics in a different way (Strickland, 2002; Hodges, 2004) and visualise and verbalise mathematics using MATLAB (Borovik, 2002). Students benefit from continuous assessment as they become more engaged with the mathematical topics. Moreover, being engaged with technology contributes to the development of students' programming skills (Stander and Eales, 2011). Introducing computerised marking of mathematics assessments benefits not only students but lecturers as well. Computerised assessment reduces the time devoted to marking and provides detailed, personalised and immediate feedback (Delius, 2004).

Alternative assessment methods include: mathematical modelling using mathematics software such as Matlab or Mathematica, group projects (e.g. web based, written assignments), presentations, projects, and portfolios (e.g. individual, group) or other (e.g. case studies (learning journal), AIM, online tutorials, R, Minitab, concept maps).

Some of the professional papers argue that projects and journals contribute to improving students' performances and to increasing their skills development (Lowndes and Berry, 2002; Lanigan, 2007). Such assessment tools build students' self-confidence and contribute to changes in their attitudes towards mathematics. CAA encourages students to revisit the problem and to find and fix their mistakes (Delius, 2004). Technology contributes to increasing the appeal of mathematics as students find the subject more thought provoking. Technology-based assessment helps students in self-evaluation. On the other hand, alternative assessment methods make mathematics lectures more interesting and help lecturers as well as students identify challenging areas early on in the semester.

2.6 Discussion

Our survey shows that most papers found to be relevant to university mathematics assessment belong to the professional papers category. This could be explained through the lecturers desire to share with their colleagues teaching practices and their interaction with various forms of technology. Theoretical papers point out drivers of change in using alternative assessment strategies. It is surprising to note the small numbers of empirical papers on investigating alternative assessment practices in undergraduate mathematics.

References

Anderson, M., Bloom, L., Muller, U. and Pedler, P. (1999) The impact of the graphics calculator on the assessment of calculus and modeling. *International Journal of Mathematics Education, Science and Technology*. 30(4), 489-498.

Berry, J. and Houston, K. (1995) Students using posters as a means of communication and assessment. *Educational Studies in Mathematics*. 29(1), 21-27.

Borovik, A. V. (2002). Implementation of the Kid Krypto concept: Computer assisted teaching of number theory and cryptography. *MSOR Connections*. 2(3), 23-25.

Burton, L. and Haines, C. (1997) Innovation in teaching and assessing mathematics at university level. *Teaching in Higher Education* 2(3), 273-294.

Cazes, C., Gueudet , G., Hersant, M., and Vandebrouck, F. (2006) Using e-exercise bases in mathematics: Case studies at university. *International Journal of Computers for Mathematical Learning*. 11, 327-350.

Chalice, D. R. (1995). How to teach a class by the modified Moore method. *The American Mathematical Monthly*. 102(4), 317-321.

Cybinski, P. and Selvanathan, S. (2005) Learning experience and learning effectiveness in undergraduate statistics: modeling performance in traditional and flexible learning environments. *Decision Sciences Journal of Innovative Education*. 3(2), 251-271.

Delius, (2004). Conservative approach to computerised marking of mathematics assignments. *MSOR Connections*. 4(3), 42-47.

Engelbrecht, J. and Harding, A. (2005a) Teaching undergraduate mathematics on the Internet. Part 1: Technologies and taxonomy. *Educational Studies in Mathematics*. 58, 235-252.

Engelbrecht, J. and Harding, A. (2005b). Teaching undergraduate mathematics on the Internet. Part 2: Attributes and possibilities. *Educational Studies in Mathematics*. 58, 253-276.

Gal, I. and Garfield, J. (1997) Curricular goals and assessment challenges in statistics education. In Gal, I. and Garfield, J. B. (Eds.), *The Assessment Challenge in Statistics Education* (pp. 1-13). IOS Press.

Garfield, J. B. (1994) Beyond testing and grading: using assessment to improve student learning. *Journal of Statistics Education* 2(1), 1-10.

Hodges, W. (2004) An experimental course in mathematical writing. *MSOR Connections*. 4(3), 29-31.

Houston, K. (2001). Assessing undergraduate mathematics students. In Derek Holton (Ed.), *The Teaching and Learning of Mathematics at University Level: An ICMI Study* (pp. 407-422). Kluwer Academic Publishers. The Netherlands.

Iannone, P. and Simpson, A. (2011). The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*. 30(4), 186-196.

Lanigan, M. (2007). Case study: Utilising learning journals in mathematics assessment. *MSOR Connections*. 6(4), 38-42.

Linn, R. L., Baker, E. L. and Dunbar, S. B. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*. 20(8), 15-21.

Lowndes, V. and Berry, S. (2002) Using projects to enable reflective learning. *MSOR Connections*. 2(3), 10-12.

Mwakapenda, W. (2003) Concept mapping and context in mathematics education. *Proceedings of the International Conference The Decidable and the Undecidable in Mathematics Education*. Brno, Czech Republic.

Oates, G. (2011) Sustaining integrated technology in undergraduate mathematics. *International Journal of Mathematical Education in Science and Technology*. 42(6), 709-721.

Pitcher, N., Goldfinch, J. and Beevers, C. (2002) Aspects of computer-based assessment in mathematics. *Active Learning in Higher Education*. 3, 159-178.

Sangwin, C. J. (2003) Assessing mathematics automatically using computer algebra and the Internet. *Teaching Mathematics and its Applications*. 23(1), 1-14.

Smith, G. and Wood, L. (2000) Assessment of learning in university mathematics. *International Journal of Mathematics Education in Science and Technology*. 31(1), 125-132.

Stander, J. and Eales, J. (2011). Using R for teaching financial mathematics and statistics. *MSOR Connections*. 11(1), 7-11.

Strickland, N. (2002) Alice interactive mathematics. *MSOR Connections*. 27-30.

Stull, J. C., Schiller, J., Varnum, S. J., Ducettte, J. and Roberts, L. (2008) The use of formative assessment in university level mathematics courses. *Journal of Research in Education*. 18, 58-67.

Part II
Case Studies of Alternative Assessment Methods

The second part of the book lists some of the alternative forms of assessment in use in institutions around the country. We approached heads of department and directors of teaching to suggest people who they knew were using assessment methods which they thought might be of interest to a wider audience. Each person nominated was approached and many kindly gave up their time to take part in a discussion about the assessment methods they were using, their reasons for using them, how they went about implementing them and what impact the new assessment had.

There were a number of factors common across groups of these case studies and we see the lecturers developing alternative methods in response to a number of different concerns. In a small number of cases, it is because of institutional constraints, but more often it is a desire to improve efficiency. Increased students numbers and other pressures on staff time mean that there is a need to streamline the assessment and feedback process. In many cases, the use of computer assisted assessment (CAA) is discussed, particularly in relation to how these systems can evaluate more complicated student responses or give more targeted feedback. In some cases concerns about plagiarism and collusion lead to changes to project and coursework systems.

Other changes have come about as staff wish to assess different skills, whether those are related directly to employment or are areas of mathematical thinking which staff do not believe the closed book examination is able to test. A number of lecturers expressed their concern about student engagement and responded to that with alternative assessment methods, perhaps with more regular, small-scale assessment or with forms of assessment which the students might find more appealing.

Of course, the case studies represent only the most opportunistic of samples and undoubtedly there are many other forms of alternative assessment in use, but the short descriptions here give examples of the solutions lecturers have developed to the problems of assessing in undergraduate mathematics degrees.

Chapter 3
Research Skills in Mathematics

Abstract This case study presents a new assessment strategy for a compulsory third year mathematics module assessed by a project. The novelty of the assessment consists of the introduction of the assessment of students' CV writing skills (written after input from the local careers office), oral presentations and peer reviewing of project drafts.

3.1 Background and rationale

This module replaced a history of mathematics module which was assessed by an individual project and a closed book exam. The department decided to replace it after students expressed dissatisfaction, feeling that its assessment was too demanding for the credits accrued. In addition, there was an institution-wide requirement to have an independent research module. The new *Research Skills in Mathematics* module has a project as final outcome, although the assessment includes presentations, CV writing and peer review of project drafts. It also introduces components aimed at improving students' employability skills.

3.2 Implementation

Research Skills in Mathematics is a year three project module spanning the whole academic year. There is only a small amount of large group, direct teaching: 6 lectures and computer labs, which includes some instruction on essay writing. Students are then assigned to supervisors according to their subject preference: pure mathematics, applied mathematics or statistics. The students meet with their supervisor three times a year for about one hour. Supervisors have a list of projects students can choose from. These include titles such as "Pierre de Fermat and his contribution to number theory", "Practical mathematical finance", "The oscillation of a liquid drop" and "History of the four colour theorem". During the first meeting the student and the supervisor decide the topic for the project. The students are also encouraged to contact the institution's careers office and discuss their first assignment: the preparation of a CV. The second assignment consists of a first draft of the project. This first draft is discussed in the second meeting with the supervisor where improvement and future development of the projects are discussed. After submitting a second draft of the project, each student reviews three of their peers' projects and the supervisor also gives feedback. In the last day of the spring term students submit the final project and they present it to their supervisor and their peers. The supervisor marks both the final project and the oral presentation.

The project is intended to be carried out independently. Students are told that their supervisors will respond to no more than 2 emails from them (in addition to the face-to-face tutorials). It is expected to be between 9 and 11 pages long and is submitted through Turnitin software to check for plagiarism. The intention is that the project should be written in such a way that a non-expert should be able to read it.

There are some perceived disadvantages. There is considerable organisation needed to co-ordinate students and supervisors (both of whom can need some extra support) and the ability to do this is very dependent on a reasonable staff-student ratio. There is also the concern that under-engaged students may get even less engaged with this type of course and assessment as there is little overt pressure on them.

The key advantages of this assessment appear to be twofold. On the one hand students engage with some transferable skills useful in the workplace, such as CV writing and oral communication, and on the other they are allowed to engage with mathematical content in a more active way. Selecting a mathematical topic with the supervisor, drafting a project and peer reviewing other students' project requires more active participation in the process of learning mathematics than a standard lecture course would.

3.3 Assessment

Stage	No. of students	Assessment pattern
Year 3	180	80% final project 5% CV writing 5% peer reviewing of projects 10% presentation

3.4 Discussion, learning and impact

Students' feedback indicates that they are happier with this module than with the previous history of mathematics module. Students on the whole appreciate being tested on a new set of skills, such as the ability to critique peers' work and essay writing. One student commented:

> I really enjoyed the Research Skills course. I think having a project group really helped as we had people to support us. For example we got together to do mock presentations. I think

> the CV assignment was a very good idea (although I moaned at the time). Have now used the CV I made for research skills to apply for both jobs and courses.

Initially, some students were uncomfortable with giving feedback on their peers' work but with practice they started to appreciate this side of the assessment as well. Students appear to gain a different appreciation of mathematics; by having to read mathematics papers and books and digest them for the project, they appear to gain a different understanding of the nature of mathematics. Some of them can become very engaged and excited by this approach. Another student commented:

> Overall though I understand that the module was important and I learnt a great deal from it.

The lecturers appreciate having one-to-one contact with their students and the fact that they get to know them better this way. Lecturers also value that students engage with mathematics in a more active and creative way, by presenting a piece of mathematics from a research paper or a book in their own words to a non-specialised audience. Marks for this module are slightly higher than they were for the history of mathematics module which it replaced. However, the assessment structure and content are completely different and the lecturer interviewed felt that any direct comparison is not possible.

Chapter 4
Students Designing Assessment Questions

Abstract This case study presents the assessment structure for a first year geometry module and a second year statistics module. This approach involves the use of the PeerWise (`http://peerwise.cs.auckland.ac.nz`) platform within the modules' assessment, allowing students to create their own question and answer sets based on the course materials.

4.1 Background and rationale

The rationale behind the use of this assessment and teaching tool is that neither of the two modules, geometry and statistics, had existing continuous assessment mechanisms and so the feedback students received during the module was seen to be limited. The lecturer came across PeerWise by chance, but thought that this tool could help address the feedback and engagement issues without involving a significant increase in staff time. The lecturer's interest in the use of this platform came from working towards his postgraduate certificate in academic practice and his interest in investigating how a new teaching tool could be used in mathematics.

4.2 Implementation

PeerWise is an open access platform that can be used for a variety of academic subjects. The software allows the students to design their own multiple choice questions together with the answers. This platform allows the lecturer to create a site for his module and restrict access to himself and his students.

The lecturer believes that providing the answers as well as designing the questions helps the students to think more deeply about the material. Other students on the same module can take the tests designed by their peers to assess their own understanding and can also leave feedback about the tests. The lecturer has access to the materials the students have produced and can monitor which parts of the module have had more questions created and which ones have been overlooked. He can also examine the general performance of the students on the questions created.

The lecturer can, if needed, intervene with comments and explanations. This platform also offers discussion forums on the mathematical concepts included in the questions. Students can offer each other valuable feedback and thus learn from each other.

The key advantages for using this form of assessment appear to be that it allows for continuous assessment and feedback without a large increase in staff time; students are encouraged to think more deeply about the material because they have to engage with it in a different way in order to write the questions and students interact with each other (through the question-taking and feedback processes). Since these key aims are generally formative, the contribution of engaging with PeerWise to the final mark tends to be nominal. For example, in the geometry module, students receive 5% for producing at least 2 questions and answers on the material and for providing feedback and comments on between 6 and 8 questions produced by their peers.

4.3 Assessment

Module	**Stage**	**No. of students**	**Assessment pattern**
Geometry	Year 1	70	45% for a course test 50% for a project 5% participation to PeerWise
Statistics	Year 2	170	20% course test 20% course test 50% open book examination 10% participation to PeerWise

4.4 Discussion, learning and impact

Prior to the use of PeerWise, there was no continuous assessment in either of the two modules and thus the students received very limited feedback. The lecturer found he had full support from the teaching committee and head of department when he suggested the use of the tool. The platform allows staff to see what students look for within the various questions posed and to assess the level of interaction with the material. Informal feedback given by the students shows that they find the use of this tool beneficial to their understanding. Moreover, high ability students seemed to be more enthusiastic and to engage more with the platform than struggling students. Their participation throughout the course and their willingness to answer and create hundreds of questions shows involvement beyond the level expected. Differences in students' performances and marks are not available yet as this is the first time the modules have been offered in such a format. Improvements for a second implementation of this assessment include the production of some sample questions (designed by the previous cohort of students).

Chapter 5
Presentations in Galois Theory

Abstract This case study presents the introduction of presentation and group work in the assessment of a year 3 Galois Theory module. The module syllabus follows a set textbook (Stewart, 2004) and as part of assessment the students are asked to present chapters from this book in group presentations.

5.1 Background and rationale

Changes in the module assessment occurred as a result of the lecturer's wish to offer students the opportunity to engage with mathematics in a more active way than the traditional lecture style. The lecturer thought it would helpful for the students to work collaboratively for a mini-project on a section of the set textbook. This assessment was also intended to encourage students to think about how to present and communicate an advanced part of pure mathematics. In the past this module was assessed by the standard combination of 90% closed book examination and 10% coursework (weekly exercise sheets). The new assessment structure retains the exam component and introduces group presentations to complement the weekly exercise sheets.

5.2 Implementation

The advantages of introducing a presentation component in the assessment of this module appear to be that students participate in more active discussion in class and engage actively in their mathematics learning with their peers. The lecturer provides some instructions in how to prepare the presentation. They also gain experience of working in groups with their peers to prepare the presentation and practice oral communication skills. It should be noted, however, that while most appreciated the course, some students did not take well to the new style, and the lecturer felt that slightly less material was covered. In addition, as the class size has grown from 10 to nearly 50, the audience feels less inclined to get involved or feels more anxious about asking questions during the student presentations. There is also a concern that, with students presenting core material, a poor presentation could affect the whole class.

5.3 Assessment

Note that for the presentations everyone in the group receives the same mark. The presentations are assessed immediately and no marks are given to the written report.

Stage	No. of students	Assessment pattern
Year 3	55	80% closed book exam 10% presentation of a small group project 10% 3 weekly exercise sheets

5.4 Discussion, learning and impact

The new format of presentations and lectures encourages students to engage more actively with the course material and gives students practice working in groups and presenting mathematics, something that the lecturer felt was missing in the final year of the course. The lecturer believes that this teaching approach produces lively class interaction with much discussion of mathematical topics amongst students. He reported that the quality of the presentations varied: some students struggled, while others presented in a very professional way. The lecturer perceives that a drawback of this assessment method is the reliability of the marks for the presentations, but as closed book examination accounts for the majority of marks for the module he does not perceive this to be a big problem. Students' feedback indicates that the module is well received. One student wrote:

> Interesting and enjoyable course in general. Having a specified text to follow also made things a lot more convenient. In addition the lecturer clarified many concepts and ideas from other algebra courses that were explained inadequately in those courses.

A few students were concerned about the novelty of teaching and assessment and indicated they preferred a more traditional type of module:

> The course was very different . . . it wasn't very effective, compared to the lecturer lecturing.

Resistance of this kind can often be caused by the novelty of the assessment (and in this case of some of the teaching) and the lack of experience students have with these assessment methods. The lecturer's experience of teaching this module is also very positive, and he is planning to retain this assessment in future years.

References

Stewart, I. (2004) *Galois Theory*. Chapman Hall/CRC Mathematics Series.

Chapter 6
Group Projects

Abstract This case study presents an example of the introduction of a group project in a first year applied mathematics module. The coursework consists of two presentations of the solution of an open-ended problem set by the lecturer and an individual written report. Marks are accrued for the mathematical content as well as the style of the presentations and the use of the typesetting software LaTeX.

6.1 Background and rationale

In this institution students seemed not to engage with applied mathematics modules as well as they did with pure mathematics, hence staff decided to design a new coursework component consisting of a group project. A group project would give students the opportunity to engage more creatively with the material while at the same time develop some transferable skills such as communication, teamwork and IT skills.

6.2 Implementation

The new assessment and structure of the module, which has only been offered for the past year, differs from the standard way modules had been taught previously in this institution. Its central idea is the group project, which includes two presentations and a report. Groups can choose between five different projects. Titles of the projects include: “Traffic lights” and “Drug concentration”. The assessment of the group project consists of:

- Presentation of the simple model (followed by feedback)
- Improvement of the model and written report
- Presentation of the improved model
- Managing the group work.

Each presentation lasts six minutes and is assessed on mathematical content, quality of explanation, structure and organisation, use of visual aids, and students’ understanding of the material. Students are required to use a wiki or Moodle to write the report on the improved model and have to use LaTeX to include mathematical expressions. Students are also required to keep a record of their group work in a wiki, e.g.

group minutes, software and literature use. Detailed guidelines on the assessment of this coursework component are posted on the institution's VLE.

Key advantages of this assessment component include the perception that by engaging with an open-ended problem, students gain insight into the creative process of doing mathematics. The presentation and group work aspect of this assessment also allow students to practise transferable skills such as group work and oral communication. However, there is an increased workload for the lecturer and, given the variability of the commitment and abilities in the class as a whole, there is likely to be some influence from the composition of the groups.

6.3 Assessment

Stage	No. of students	Assessment pattern
Year 1	170	90% closed book exam 10% coursework (5% written report; 5% presentations)

6.4 Discussion, learning and impact

The lecturer believes there is more engagement and enthusiasm among students. He also welcomes the opportunity to engage the students in a less formulaic and repetitive approach to applied mathematics by setting open-ended problems for the group projects. Students' feedback provided at the end of the semester shows their satisfaction with the new assessment. However, the lecturer does not consider that changes in students' performances can be measured in terms of grades for this module, as the syllabus and assessment have changed radically from previous years and a direct comparison is not possible. The drawback of this type of assessment is that it is time consuming for the lecturers involved, but in light of students' satisfaction and engagement the department will offer the module in this format again.

Chapter 7
Online Quizzes

Abstract This case study presents the use of weekly online quizzes as a coursework component for a year 2 linear algebra module.

7.1 Background and rationale

Changes in the assessment structure of this module came from the lecturer's desire to help students engage with the content from the start, particularly in the light of the definition-heavy structure of this part of mathematics. It was felt that assessment which engages students on a weekly basis can help them understand the material and keep in touch with what is covered during the lectures. Robust and efficient online assessment also has the potential to contribute reducing the lecturer's workload and allows him to have a better understanding of the topics with which students are struggling.

7.2 Implementation

Traditionally, the coursework for this module consisted of weekly paper-based exercise sheets which were then marked by the lecturer or by postgraduate students. Last year the exercise sheets were replaced by online weekly quizzes using Moodle. The system presented benefits for both staff and students. On the one hand, with support from the IT services, staff can generate fairly complex questions and in the long term build a large database of problems. The basic system in use in the department allows the lecturer to set up multiple choice questions as well as numerical questions, but with some extra work questions generated in Maple can be imported into the system to allow for quite complicated questions to be used.

The system also appears to work well for students who appreciate the opportunity to complete this coursework in their own time and online. These quizzes are based in Moodle, a VLE which also provides a platform for students to access and see course materials.

The key advantage of this assessment method is students' continuous engagement with the material. In this way students can come to terms with the basic definitions, calculations and concepts of this part of mathematics by practising with the online quizzes on a weekly basis. Continuous assessment also prevents students from adopting a last minute revision strategy before the final exam. It also significantly reduces the marking and administration time for the lecturer.

7.3 Assessment

Stage	No. of students	Assessment pattern
Year 2	200	90% closed book exam 10% coursework: weekly online quizzes

7.4 Discussion, learning and impact

Since the introduction of online weekly quizzes, exam marks increased slightly compared to previous years, but not significantly. Students apparently welcomed the opportunity to do something different and enjoyed doing quizzes in Moodle. The uptake of such quizzes every week is higher than the submission of weekly exercise sheets in previous years. Particularly in a course like linear algebra, there is a perceived need to work hard to maintain student engagement and the online quizzes seem to have done this.

Students also appreciate the opportunity of receiving instant feedback rather than having to wait for the exercise sheets to be marked. The lecturer tailors the quizzes to the material covered in the lectures during the week and feels that this helps students' understanding of notes better and reinforces important points that could be otherwise overlooked. This system has also reduced the marking load for the lecturer while at the same time allowing him to have a weekly picture of students' progress. There is a need for technical support and know-how to implement the quizzes and that requires a level of departmental resource. Amongst the other drawbacks of using this system is the feeling that it is not possible to test conceptual understanding with this sort of quiz, but as this is only one small part of the module assessment this is not perceived to be a big problem.

Chapter 8
Presentations of Applications of Pure Mathematics

Abstract This case study presents the use of individual posters and presentations as coursework for a year 3 Game Theory module. The originality of the assessment in this module consists of creating a poster which illustrates an application of game theory to real world problems and presenting it.

8.1 Background and rationale

The introduction of individual posters and presentations as part of assessment of this module was motivated by discussion in the department about enhancing students' employability skills. There is a concern in the department about communication skills: students are offered only three opportunities to undertake a presentation in modules from the department, of which this module may be the first opportunity. In the lecturer's experience this sequence of assessment by presentations really improves students' communication skills: by the time they come to the presentation associated with their final year project, they appear to have grown in confidence.

8.2 Implementation

The Game Theory module was previously assessed by one-third class tests and two-thirds final exam. The coursework now consists of a poster and a 3-minute presentation. Students can choose among a list of small projects such as

- Contribution of John Nash to game theory
- The use of game theory to describe sexual selection
- Is believing in God a game theoretical problem? Consider Pascal's wager

as the subject for their poster. Students are also given the option to work as a group, but the final product has to be an individual contribution. The module outline contains a list of transferable skills developed by this assessment which includes: independent research, presenting results succinctly both on the poster and orally and group working.

The key advantages of this assessment are that the nature of the projects helps students appreciate how mathematics can be used to solve real-life problems. It also helps them develop both their oral and written presentation skills. However, there is a concern that some students can become extremely nervous about the presentation component which may cause issues about fairness and equality.

8.3 Assessment

Stage	No. of students	Assessment pattern
Year 3/4	55	67% closed book exam 33% posters and presentations

8.4 Discussion, learning and impact

The lecturer who coordinates the Game Theory module believes that the coursework format she has adopted will contribute to an improvement in students' performance in other modules in mathematics. Moreover, students gain experience in new ways of presenting both orally and in written form. In her experience, the oral presentations have given the shyer student confidence to speak in public and present their own ideas and work on the topic of the poster. Her view is that the majority of students enjoy the process, engage with it and are pleased with the outcomes. It also gives students new assessment experiences: some have never made a poster before, never researched a story, or worked as part of a team. This assessment also gives the lecturer the opportunity to get to know her students better and to see who might be suitable for postgraduate studies.

Chapter 9
Continuous Assessment in a History of Mathematics Module

Abstract This case study presents the assessment structure for a third year history of mathematics module. Assessment includes a variety of methods such as essay writing, peer-assessed posters for mini-projects and the solution of a mathematical question with the appropriate historical tools. The course aims to improve students' understanding of mathematics as a product of history and culture as well as to improve their essay-writing and communication skills.

9.1 Background and rationale

This module is team-taught by three lecturers. They designed the assessment schedule for this module both because of their interest in the history of mathematics and essay-writing and because they felt that students would welcome the opportunity to experience a variety of assessment methods in their third year. The lecturers felt that in this way the students would engage with some mathematical content, but also would practise essay writing, team working and presenting ideas through posters. These skills, they feel, are important in the workplace and are not practised enough in mathematics degrees.

9.2 Implementation

The module in its current form is divided in three parts - ancient Greek mathematics, the development of calculus and the history of statistics - each taught and assessed by different lecturers. For each part of the module students work on the historical development of a branch of mathematics. For example, the students will study Cauchy's development of a rigorous basis for calculus and Galois' work in algebra. There is a wide variety of assessments in the module. Students undertake two essays which focus on a historical topic, but are also expected to include some mathematics. They also work in small groups (of around 5 students) to prepare a poster on a topic and to peer-assess other groups' posters on visual impact, clarity, scope and use of references. One of the lecturers assesses his component by setting traditional-looking mathematics questions which require the use of the mathematics of the day: for example, requiring students to solve a calculus-type problem using the methods Huygens developed before Newton and Leibniz.

The key advantages of these forms of assessment appear to be that students practise transferable skills such as essay writing and communication, as well as engage critically with the historical development of mathematics. The main lecturer believes that essay writing is a very important skill for students to have, independent of what job they will take after graduation. He also feels that the nature of these forms of assessment means that all students can engage: sometimes in an examination, it was felt that some students simply could not do anything and failed very badly.

However, with this assessment scheme the students sometimes raise concerns about the group work being awarded a single mark for everyone in the team when they perceive that not everyone contributed equally. Some also feel that there is too much assessment in one module.

9.3 Assessment

Stage	No. of students	Assessment pattern
Year 3	55	40% essay 60% coursework (20% essay, 20% peer-assessed group poster, 20% mathematical problem)

9.4 Discussion, learning and impact

Staff noticed substantial improvement between the first and the second essay, in terms of how students conceptualised the written work, on the reference style employed and on their ability to focus on the topic. The main lecturer reports that there was a significant increase in marks and a mark distribution that is not typical of mathematics modules. Marks last year were gathered at the top and the bottom of the spectrum; many students achieved high marks and there were no fails. The lecturer believes that for mathematics this pattern of marks is unusual. However as this is not a typical module in mathematics, this unusual pattern was not perceived to be a significant problem. The lecturer also describes how students engaged with this module and believes that many did not do enough to read independently around the topic for the essays, which he would have expected them to do. He believes that this is because in mathematics independent study has a rather different shape than in history and students were not used to this. On the whole the lecturer is satisfied about how the module ran and plans to offer it in the coming years in the same form.

Chapter 10
CAA with a Flexible Mark Scheme

Abstract This case study presents the use of an algorithmic e-assessment for summative and formative assessments in two first year modules. This automated process simulates the marking and feedback mechanisms used by humans, in particular giving method marks and marks for partially correct answers, and the system can be used to re-mark according to particular conditions.

10.1 Background and rationale

The university has used e-assessment for more than 16 years both for summative and formative assessment. However, it experienced difficulties with the robustness of the commercially available systems and with the lack of flexibility in the question design. For example, the system in use was unable to assess questions which had an infinite number of correct answers or identify whether students had undertaken some steps of a calculation correctly. Frequent crashes in the e-assessment system proved detrimental to students' and academics' experiences. The lecturer interviewed has a strong background in computing and decided to design his own e-assessment system. He designed a new system which is robust and fully algorithmic in both the question generation and question marking stage. For example, the system is designed to easily accommodate *continuation marking*. That is, in the marking of a question with multiple parts, the system will detect cases where a student's incorrect answer may be deemed partially correct based on one of their previous answers. Another example of the algorithmic marking process is the case of *verification marking* whereby the marking algorithm can perform verification tests on the student's input to determine whether the solution is correct. This is particularly useful for questions for which there are many acceptable solutions. For example, if the question asks the student to find a vector that is orthogonal to a given plane, the system will perform the verification that the student's input is a non-zero scalar multiple of a sample solution.

The system generates immediate feedback and marks and does this efficiently. The presentation of this feedback and marks may be given to the students immediately following the assessment attempt or it may be delayed until after the end of the assessment period. The algorithmic questions can detect students' use of mal-rules and report such incidents in the feedback.

10.2 Implementation

This system is now used in 17 modules across the department, including in linear algebra and in mathematics for engineers. The system is used for both formative and summative assessment. For summative assessment students can take the tests in their own time, and the system then marks their tests. In this mode students enter an answer to a part of the question (typically questions have 4 parts) and receive a mark from which they can see whether their answer is correct. Students have two attempts for each assessment. At the end of the assessment attempt the students receive feedback. An important feature of the system is that if the lecturer feels that there is a glitch with the marks or that the students have used a procedure which is partially correct but obtained the wrong answer, the lecturer can, at the end of the assessment period, adjust the mark scheme and ask the system to re-mark all the assessments.

The key advantage of such a system is that it can be used for continuous formative assessment as well as summative assessment. It also saves staff time when marking summative tests. Amongst the drawbacks is that the summative assessment is done in the students' own time so there is no control over plagiarism or collusion. On the other hand the lecturer stressed that this is the case for any piece of assessed work not produced under exam conditions.

10.3 Assessment

Module	Stage	No. of students	Assessment pattern
Linear Algebra	Year 1	120	70% closed book exam 30% 2 computer based tests in non-controlled conditions
Mathematics for Engineers	Year 1	260	20% 4 computer-based tests 80% two closed book exams

10.4 Discussion, learning and impact

Some of the benefits include timing and management of a student's assessment. Academics can write their own algorithmic questions as the process of programming a question in this system is straightforward. Students receive prompt and detailed feedback on their submissions. Upon implementing the new systems of e-assessments, both in summative and formative form with weekly tests, the pass rate for the Mathematics for Engineers increased from 65% to 85%, which was attributed to the students regularly engaging with the material through the weekly tests. Student satisfaction with these assessments is very high. Last year 70% of the responding Mathematics for Engineers students said they were “very useful”, 17% “quite useful or ok” with only 9% stating they didn’t use it.

The system is now in use in many modules and the department is fully supportive of the further development of its features.

Chapter 11
Portfolios using Maple

Abstract This case study presents the assessment strategy for a year 1 computational mathematics module. Assessment consists of a portfolio of questions and a mathematical modeling project developed using the mathematical software package Maple.

11.1 Background and rationale

Changes in the assessment structure for this module were introduced following the university's drive to improve students' employability skills. The approach to assessment used in this module came from the lecturer's desire to improve students' programming skills, to draw together concepts from across the curriculum, to develop students mathematical modelling abilities and to help them use their mathematical knowledge across different subjects.

11.2 Implementation

This is a year-long module divided in two parts: computational mathematics and mathematical modelling. The first part is assessed by a portfolio of weekly questions students are assigned from a set list, linked to the mathematical topic taught in that week. An example of the question in the portfolio is:

> **Maple Tutorial Sheet 7**
> Use the `dsolve` command to solve the differential equation
>
> $$\frac{1}{x}\frac{dy}{dx} - \frac{2y}{x^2} = x\cos x.$$
>
> Plot the solution with constant of integration equal to -1, -0.5, 0, 0.5, 1 on the same graph.

The emphases of the questions are on understanding mathematics, but also on learning the use of Maple.

The second part is assessed by a group project and presentation in mathematical modelling. This is done as a group, with the grade given for a mixture of assessment tasks including presentations and report writing, and the module contains teaching focused specifically on developing these skills.

Students are required to use Maple for both assessments and they are taught how to programme in Maple in the first part of the course. Developing computer programming skills and developing effective contribution to group work are among the set learning outcomes of this module.

Key advantages of this assessment schedule are the focus on employability skills including programming and computer literacy, communication and writing. It is also seen to have benefits in developing many of the skills needed for the more substantial final year project. As with any form of assessment, one concern is that there are a small number of students who do not engage or put in sufficient effort.

11.3 Assessment

Stage	No. of students	Assessment pattern
Year 1	78	60% mathematical modelling group project and presentation 40% portfolio of questions

11.4 Discussion, learning and impact

The lecturer coordinating this module describes how the university drive to include employability skills in the teaching of mathematics has partly motivated their choice of assessment. Being able to programme and being computer literate are skills that any employer will value and that students do not have at the start of their degree. Marks so far for this module are high, near 60-70%, showing that most students put in the time and effort required. The lecturer also appreciates seeing students grow in confidence using the software as the year progresses. The drawback of this assessment schedule is the heavy marking workload, especially in view of the growing number of students joining mathematics degrees.

Chapter 12
Group Projects with Individual Presentations

Abstract This case study presents the assessment of a third year project module. The assessment consists of combining a written *group* project with *individual* presentations of the same project.

12.1 Background and rationale

Before the introduction of the new form of assessment, the final year project had been done on an individual basis. The change to a group project was motivated by the workload that came with the need to supervise an increasing number of students in weekly one-to-one sessions. After a trial of assessing the groups for both the written output and the group presentations, staff decided to retain the group assessment for the written work but assess the students individually for presentations on aspects of the project. The idea of combining this group project with an individual oral presentation was motivated by concerns that individual contributions were not being suitably recognised. Group projects were also seen as a way to make projects more stimulating for both staff and students. Asking students to work in groups would also develop their employability skills.

12.2 Implementation

Groups of 4 to 6 students are assigned to supervisors and work with them towards the production of the report for the group projects. They meet weekly with the assigned supervisor, and produce a 40-page dissertation on the chosen topic. For the assessment of the individual presentations, each member of the group selects an aspect of the project and presents it orally for 10 minutes in front of a panel consisting of 3 members of staff including the supervisor. There are also an additional 5 minutes for questions from the panel to test their understanding.

Examples of projects are:

The shape of space Surfaces occur all around us in the real world. The idea of planes and spheres is familiar from an early age, but so too are more complicated surfaces such as the annulus (the shape of a ring doughnut). Is it possible to classify all possible surfaces? Surfaces can be much stranger than we might initially expect. The Möbius strip can be formed out of a piece of paper but only has one side, while the Klein bottle, although closed, does not separate its ‘inside’ from its ‘outside’. The aim of this project is to introduce the area of maths which studies surfaces, called topology.

Barrier Options Under certain simplifying assumptions, the valuation of financial options on the stock market can be carried out by solving a partial differential equation known as the Black-Scholes equation. The main objective of this project is to obtain solutions of this equation for barrier options, where the asset price is constrained to lie between upper and lower bounds. Using techniques for solving partial differential equations, including coordinate transformation, Fourier series and separation of variables, it is proposed to obtain an analytical solution of the problem. Properties of this solution will then be examined to determine how the option value depends on the various financial parameters involved.

Key advantages of this assessment schedule are that students experience working in groups and presenting mathematics to an audience. The projects also foster independent thinking and communication skills. From the staff perspective, working with a group means that ideas need only be explained once, the supervisions can be more interactive and a good supervisor can foster a lively discussion between themselves and the group of students.

12.3 Assessment

Stage	No. of students	Assessment pattern
Year 3	90	50% written group project - group mark 50% individual presentation of the group project

12.4 Discussion, learning and impact

The lecturer believes that the new assessment structure for this module allows for a more realistic and finer assessment of the students' performance. The projects topics are changed every year and they are proposed by the staff involved with supervising the groups. The lecturer believes that this helps maintain staff interest in the projects. The main drawback of this assessment method had been the difficulty in assessing each student's individual contribution, but the individual presentations have now addressed this issue. The lecturer also reports a few instances where the groups of students were not functional because of personality clashes between members and adjustments had to be made by changing the composition of some of the groups.

Chapter 13
Presentations and Quick Quizzes

Abstract The case study presents a new approach to assessment for first year calculus. The aim of this module is to bridge school to university mathematics as well as use presentations. The lecturer has introduced fortnightly quizzes on basic material from the course.

13.1 Background and rationale

Changes were made because the university restructured the credit scheme of its undergraduate degrees. Together with institutionally driven reasons for change, staff felt that the feedback usually associated with fortnightly exercise sheets could be improved and wished to offer students better and quicker feedback. In addition, the standard coursework (which had consisted of exercise sheets each week) did not seem a very good predictor of final marks: students often got very high marks, but there were concerns about copying. The main aim of this module is to ease the transition from high school to university mathematics and in its previous form this module had no final exam. The evidence appeared to suggest that students did not prepare adequately and often performed badly.

13.2 Implementation

As well as traditional lectures, students are taught in small tutorial groups of about five or six, and this is aimed at helping students settle in the course. Students were expected to attend tutorials and this improved their engagement and preparation of the material. Quizzes were added to the coursework about three years ago and are entirely based on definitions and factual recall. For each quiz, students have to solve 5 questions within 5 minutes and are marked either 0 or 1.

An example of questions included in the quizzes is:

1. Write down an expression for $(a+b)^n$ given by the binomial theorem, using the Σ notation.
2. Given polar coordinates r and Θ, write down equations for the corresponding Cartesian coordinates x and y.
3. Write down a formula for the root mean square of the function $y = f(x)$ between $x = a$ and $x = b$.

Students also receive general feedback at the beginning of the second semester after the test they take in January which encompasses discussion of the most common mistakes. Examples of the feedback are:

> Most used the correct partial fractions. Fewer students than usual for such questions integrated the quadratic term as a log, but some did not understand the effect of the 3 in $1+3x$.

and

> Generally, lots of students continue to give decimal approximations to answers (as well as the exact form). I did not deduct marks, but you should not do this!

In addition to these quizzes, presentations were added to the assessment at the beginning of the current academic year. The purpose of these presentations was to make students interact and become more engaged with mathematics. Lecturers believe that the presentations highlighted some misunderstanding of fundamental mathematics concepts which were not apparent from other types of assessment. The presentations are held during the tutorial sessions.

A key advantage of this assessment is students' continuous engagement with the material. In addition, the students' performance on the quizzes can be very low; at one point the median quiz score was zero. This can bring them to the realisation that they lack understanding of some very basic mathematical concepts and push them to commit basic material and definitions to memory. While there was a concern that the move away from written homework might impact on their learning to write mathematics correctly, this does not appear to have become a significant problem.

13.3 Assessment

Stage	**No. of students**	**Assessment pattern**
Year 1	130	80% closed book exam 15% 4 quizzes taken over two semesters and 4 traditional exercise sheets 5% January test

13.4 Discussion, learning and impact

In the past while students agreed the material covered in this module is easy, their course marks were low. The new assessment structure helps students realise they lacked certain mathematics skills and basic knowledge. The introduction of quizzes and presentations had a positive effect on final exam marks with lower quartile marks improved: the median mark for the last cohort was 35%, whilst at the end of the current academic year it was 40-45%. The lecturer also noticed improvement in students' engagement with the course. Students responded positively to the experience of giving a 10 minute presentation. Staff members feel that this assessment structure will be maintained, as students prepare better and engage better with mathematics. The quizzes and presentations also gave them a good understanding of gaps in their knowledge and misunderstandings of material. Amongst the drawbacks is the requirement on staff time for marking, but the lecturer feels that the advantages of this assessment structure outweigh the disadvantages in terms of workload. Another concern raised by the lecturer is the assessment of the presentations as several members of staff are involved in the assessment and marks seem not to be consistent. The department is now moving to increased use of class tests, given in a relatively formal way, but with quite small proportions of the final mark dependent on them as they feel this approach increases students' work on the material during the year and improves their preparation.

Chapter 14
Mini Projects and Library Tasks

Abstract This case study presents the assessment strategy of a third year module called *Information Skills in Mathematics*. The assessment consists of using three distinct tasks – including a mini project – aimed at enhancing students' research, communication and presentation skills.

14.1 Background and rationale

Information Skills in Mathematics is a module assessed entirely by coursework, which has run for 6 years. It was developed in response to an external review which recommended that all students undertake at least one project in their undergraduate year. Aimed at enhancing students' employability skills, the module it is now assessed by three separate tasks. The tasks assess not only mathematical content, but also transferable skills such as mathematics word processing, independent research, and written and oral communication skills.

14.2 Implementation

The module is divided in three parts. First, students learn $\LaTeX$ and have to submit a $\LaTeX$ assignment. The second part covers library skills: for this, students select a topic from a given extensive list and are required to find 2 books, 3 journals articles and 4 web resources relevant to the chosen topic. They are then asked to evaluate the sources and write an essay no longer than 2000 words. This task teaches students how to do a literature search on a mathematics topic and how to use the library resources appropriately. The third task is the mini-project. This consists of reading and reporting their understanding of one research article in mathematics. Each student is assigned a different research article (there are 130 different articles on the list). Examples of such articles are:

> *Multiplicative groups of singular matrices* B. Lang and H. Liebeck, Mathematical Gazette, (60), 1976, 38-47.
>
> *Abstract:* In a group every element must have an inverse. So it would seem at first sight that in a group of square matrices, all the matrices must be invertible, i.e. they must be non-singular (with non-zero determinant). However, this isn't so.
>
> *The two-state Markov process and additional events* L.Rade, American Mathematical

Monthly, (83), 1976, 354-356.

Abstract: This is the famous method of interpreting the Laplace transformation in Probability, on one of the simplest models of continuous time Markov chains.

For the mini project each student is assigned to a supervisor (the member of staff who suggested the paper in the given list) and they meet three times during the semester for between half an hour and an hour each time. The final report is expected to be 5 to 10 pages long and the students present the project in a 5 minute presentation. The student's supervisor assesses the mini projects. The projects are submitted through the Turnitin software which helps detect plagiarism.

The key advantages of this assessment schedule is that it allows students to engage with a piece of research in mathematics with the final mini project while at the same time it helps them develop employability skills such as written and oral communication, synthesising and evaluating sources and specialised word processing. The possibility has been discussed of replacing some parts of the LaTeX section of the course with an aspect of industrial mathematics in future.

14.3 Assessment

Stage	**No. of students**	**Assessment pattern**
Year 3	100	15% LaTeX assignment 25% library assessment 60% mini project and presentation

14.4 Discussion, learning and impact

On the whole, students engage enthusiastically with this module and its assessment structure, although the lecturer interviewed believes that many students see it as an easy module. The lecturer also reports that the marks students gain on this module tend to be higher than for other mathematics modules. This may be because supervisors are too lenient when marking the mini projects and the presentations and do not take sufficient account of the fact that some mathematics topics suggested for the projects are more complex than others. The difficulty in assessing the written mini project has been partially overcome by the presentation which gives the assessors a better indication of the students' level of understanding. The drawback of this assessment schedule is the complexity of its administration. Coordinating many supervisors and a cohort of 100 students is very time consuming. Similarly, there is a heavy assessment load: the mark comes from averaging 2-3 different evaluations of each project.

Chapter 15
CAA with Natural Input Syntax

Abstract This case study presents the implementation of a computer assisted assessment system (CAA) for a first year module in algebra and calculus which interprets student input and generates more appropriate feedback. This system was developed eight years ago by a lecturer in the department and a colleague in another institution. It has been used since in the School of Mathematics for year one modules, in the past two years for mathematics modules in the School of Physics and is being used in other institutions in the UK and internationally.

15.1 Background and rationale

The lecturer who developed this system has a personal interest in designing and using CAA. In addition, due to the increase in class sizes in the past few years, the marking load for year one modules had become unmanageable. The new method reduced the amount of time and resource associated with marking as well as allowing the lecturers to gain useful information about their students' performance. In addition, at the time it was developed, there were concerns that existing CAA systems did not do a good job of interpreting more complicated student input or of providing useful feedback.

15.2 Implementation

This CAA system uses a computer algebra platform called Maxima. It includes analysis tools which support the lecturer in tracking both individual student's answers and the cohort's answers to each question. The system can generate similar but distinct questions for each student in a pseudo-random way so that students can still discuss types of questions together without lecturers being worried about plagiarism. These features set the system apart from other online assessment systems that simply select answers from a list of predetermined multiple choice or multiple response questions. Rather than expecting a fixed response, the system can allow students to enter their responses in a syntax similar to a programming language or graphics calculator and can interpret whether the input is equivalent to the expected answer. This CAA system provides individualised feedback, assigns a mark, stores outcomes in a database, and creates a profile for each student.

Figure 15.1 gives an example of a question and feedback to an incorrect solution.

Give an example of a function $f(x)$ with a stationary point at $x = 5$ and which is continuous but not differentiable at $x = 0$.

$f(x) =$ x*(x-10)

Your answer was interpreted as:

$$x \cdot (x - 10)$$

Your answer is partially correct Your answer is differentiable at $x = 0$ but should not be. Consider using $|x|$, with is entered as `abs(x)`, somewhere in your answer.
Your mark for this attempt is 0.67.
With penalties and previous attempts this gives 0.67 out of 1.

Fig. 15.1 Example question and feedback.

In addition to the input and feedback systems, the key benefits of this CAA system include the possibility of creating similar but distinct questions for students, its ability to award partial credit and the ability of local users of the system to modify the system and author questions. However, the developer acknowledged that the amount of time needed both to develop the system and to author questions can be significant.

15.3 Assessment

Stage	No. of students	Assessment pattern
Year 1	230	80% closed book examination 10% CAA weekly quizzes 10% exercise sheets

15.4 Discussion, learning and impact

Students appear to enjoy the mathematical sophistication of the system, the rapid and comprehensive feedback received, the opportunity to have multiple attempts for a mathematics problem, and the system's allocation of marks. The lecturer also believes that the fact that answers need to be input in the CAA system using a more natural syntax is an added benefit to the students. Amongst the drawbacks is that in its current form this CAA system only assesses the final answers and not the process of finding this answer. This is an area which might be looked into as a future development with the possibility in the longer term to enable the system to assess the validity of students' proof.

Chapter 16
Assessment in a Moore Method Module

Abstract This case study presents the assessment of a year one module taught with the Moore Method and entirely based on problem solving. Assessment is divided between participation in class activities and written work.

16.1 Background and rationale

There was a feeling amongst staff that students were not engaging in "doing" mathematics, but that they were passive and learned in a very procedural way. This module was introduced about 7 years ago to address this problem. The change in the teaching method led to a different assessment regime.

16.2 Implementation

The course follows some of the general principles of the Moore Method: the lecturer does not impose solutions; rather students have to find answers to given problems without the help of supporting material such as books. It forms half of a module (along with a semester on the impact of mathematics). The module runs with relatively small numbers: it began capped at 20 students, but is now two groups of 20. The group size allows for a more discursive and interactive form of teaching. It is an optional module in the first year and can be taken by students from other departments. Examples of the problems given during the module are:

> A ladder stands on the floor and against a wall. It slides along the floor and down the wall. If a cat is sitting in the middle of the ladder, what curve does it move along?
>
> A circle and a point A inside it are given. Points B and D lie on the circle. Find the set of vertices of the rectangles ABCD.

Students are expected to work independently to solve the problems and present their solutions in class; the solution is then subject to discussion in the seminar sessions. The lecturer assesses the presentation of the solution to the seminar as well. These presentations are videoed and assessed later. The hand-written proofs submitted as part of this module are also assessed. These proofs are relatively strictly marked: the emphasis is on correctness and students receive marks only if the proof is perfectly correct. On the other hand, students are permitted to rewrite their solutions as often as they like.

The key advantage of this assessment is that students gain insight into the process of doing mathematics independently and in an active way. They also practise proof writing. The course can be quite intensive and some students do not rise to the challenges set by putting in sufficient effort. Some students can find presenting their solutions in front of their peers daunting and the lecturer admits that he sets high expectations about how prepared they need to be for each session. The module is also quite resource heavy in terms of teaching.

16.3 Assessment

Stage	**No. of students**	**Assessment pattern**
Year 1	40	50% contribution in class 50% submitted solutions to problems

16.4 Discussion, learning and impact

The lecturer enjoys this method of teaching as it provides the opportunity to get to know the students, and to see their engagement with, and understanding of, mathematics growing throughout the term. While it is very demanding to write the problems for this module each year, the lecturer sees this as a very satisfying challenge. During this module students come to realise how much time it takes to solve a problem and they get a good sense of what mathematics is about. Many also become quite competitive and much more motivated when they present in front of their peers. The lecturer notes that this module seems to attract students who subsequently pursue postgraduate studies in mathematics, and this is again very satisfying for staff involved.

Part III
Assessment Projects

The final part of the book looks deeper at the evaluation of some alternative assessment methods. Academics were asked to bid for resources from the MU-MAP Project to either trial a new assessment method or to evaluate the impact of an existing one. In some cases, people chose to focus on students' views of the assessment. Some chose to look at the perception of lecturers who are using particular methods. Still others looked at the practicalities of implementing an alternative assessment system and evaluate the advantages and disadvantages compared to the methods they replaced.

Chapter 17
Summative Peer Assessment of Undergraduate Calculus using Adaptive Comparative Judgement

Ian Jones and Lara Alcock

Abstract Adaptive Comparative Judgement (ACJ) is a method for assessing evidence of student learning that is based on expert judgement rather than mark schemes. Assessors are presented with pairs of students' work and asked to decide, for each pair, which student has demonstrated the greater proficiency in the domain of interest. The outcomes of many pairings are then used to construct a scaled rank order of students. Two aspects of ACJ are of interest here: it is well suited to assessing creativity and sustained reasoning, and has potential as a peer-assessment tool. We tested ACJ for the case of summative assessment of first year undergraduates' conceptual understanding of a specially designed calculus question. We report on the relative performance of peer and expert groups of assessors, and the features of student work that appear to have influenced them. We consider the implications of our findings for assessment innovation in undergraduate mathematics.

17.1 Introduction

This project involved implementing and evaluating an innovative assessment of undergraduate calculus students' conceptual understanding of properties of two-variable functions. The innovation replaced a traditional computer-based test and contributed 5% of each student's grade for a first year calculus module. It comprised two parts. First, students completed a written test designed to assess conceptual understanding that was specially designed for the innovation, shown in Figure 17.1. Second, students assessed other's responses to the test online using an Adaptive Comparative Judgement (ACJ) method.

ACJ is an approach to assessing student learning that is based on holistic judgements of work rather than aggregated item scores (Pollitt, 2012). As such it offers promise for assessing conceptual understanding and for use as a peer assessment tool. It has been demonstrated to be effective in a variety of settings, from technol-

Ian Jones, e-mail: `i.jones@lboro.ac.uk`
Lara Alcock, e-mail: `l.j.alcock@lboro.ac.uk`
Mathematics Education Centre
Loughborough University

Conceptual Test Question

Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by:

$$f(x,y) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } x \geq 0 \text{ and } y \geq 0 \\ -x & \text{if } x \geq 0 \text{ and } y < 0 \end{cases}$$

Describe the properties of this function in terms of limits, continuity and partial derivatives. You should explain and justify your answers, and you may do so both formally and informally, using any combination of words, symbols and diagrams.

Fig. 17.1 Written test question designed to assess conceptual understanding

ogy teacher training (Seery, Canty and Phelan, 2011) to GCSE mathematics (Jones, Swan and Pollitt, in progress).

ACJ is derived from a well-established psychophysical principle (Thurstone, 1927) that people are far more reliable when comparing one thing with another than when making absolute judgements. Assessors are presented with pairs of scripts and asked to decide which student is the more able mathematician. The judgements of many such pairings are then used to construct a final rank order. This is usually done using a Rasch model which produces residuals for each judgement, thereby allowing the linearity and coherence of the final rank order to be explored in detail.

Until recently comparative judgement was not viable for educational assessment because it is tedious and inefficient. The number of required judgements for producing a rank order of n scripts would be $\frac{(n^2-n)}{2}$, meaning that for the 168 scripts considered here, just over 14000 judgements would be needed. However the development of an adaptive algorithm for intelligently pairing scripts means the number of required judgements has been slashed from $\frac{(n^2-n)}{2}$ to $5n$, so that 168 scripts now require only 840 judgements.

17.2 Implementation

17.2.1 Test design and administration

The written test was developed by the course lecturer (the second author) specially for this project. We considered various practicalities when deciding on the precise test structure and administration. First, timing of the test in relation to the course meant that students had been provided with definition-based lectures and exercises related to the concepts of limits, continuity and partial derivatives for functions of two variables, but that they had done only minimal work on the last of these. Second, we wanted to ask a question that would prompt students to think more deeply about these concepts, which are known to challenge students in different ways and to different extents (Pinto and Tall, 2001), than would routine exercises or even variants of routine proofs: such routine work is required in other tests within the module. Third, we wanted an individual written test in order to fit with the requirements of the ACJ system but, because it would replace an online test, we did not want something that would take up a lot of lecture time.

As a result, we decided to set the test question given in Figure 17.1, which we hoped would allow students considerable flexibility in choosing how to respond, and which would prompt them to think about whether and how concepts from the course applied in a non-standard situation. In order to encourage this thinking without taking up excessive lecture time, we distributed copies of the question to the students six days in advance of the lecture in which the written test was to take place. The test was administered in a lecture under exam conditions: students were allowed 15 minutes to complete the test and were told that their answer must fit on the single side of A4 paper as provided.

17.2.2 Peer use of ACJ

33 students opted out of their scripts being used for research purposes and we discuss only the remaining 168 scripts in the report. The scripts were anonymised by removing the cover sheet, and then scanned and uploaded via a secure file transfer protocol to the ACJ website[1].

The day after the written test, a researcher explained the paired judgements activity and demonstrated the ACJ website to the students in a lecture. The researcher told the students that they would log in and be presented with 20 pairs of scripts, and that they should decide, for each pair, which author had demonstrated the better conceptual understanding of the question. A screenshot of the user interface is shown in Figure 17.2. He advised them that each judgement should take on average around three minutes and that the total work should take no more than one hour.

A user guide was provided on the course VLE page to support students with technical aspects of ACJ, and drop-in support sessions were offered in a computer lab during the exercise. In practice, no technical problems were reported and the only help requested were password reminders.

17.2.3 Rank order construction

Once the students had completed the online judging we constructed a rank order of scripts by fitting the judgements to a Rasch model (Bond and Fox, 2007). The outcome of the Rasch analysis was a scaled rank order. Each script was assigned a parameter value and standard error along a logistic curve. The final rank order of scripts produced by the students is shown in Figure 17.3.

[1] The ACJ website is called "e-scape" and is owned and managed by TAG Developments, the e-assessment division of Sherston Software.

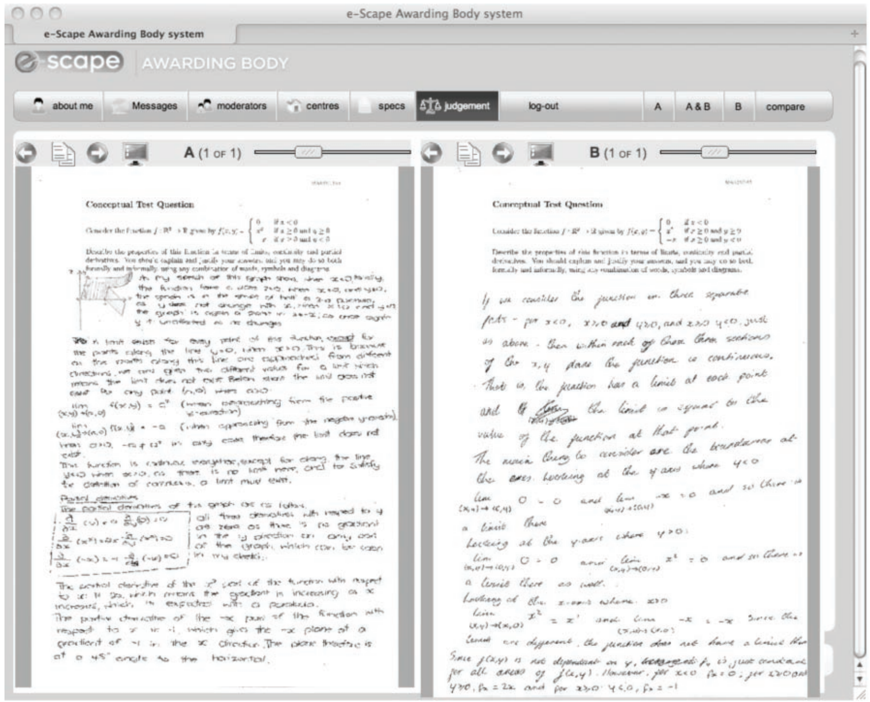


Fig. 17.2 The "e-scape" system's ACJ user interface.

Rasch analysis produces a host of measures that can be used to explore the stability of the rank order. A key measure is the internal consistency, analogous to Cronbach's α, which can be considered the extent to which the students' judgements are consistent with one another. The internal consistency of the students' rank order was .91, an acceptably high figure.

17.2.4 Allocation of grades

A rank order produced by ACJ can be used to allocate grades to students in the standard way. This can be done using norm referencing, for example, allocating the top 20% of scripts a grade 'A' and so on. Alternatively it can be done using criterion referencing. This requires sampling scripts from across the rank order and comparing them against agreed assessment criteria. Boundary scripts within the rank order can then be identified and grades applied accordingly. In our case the students will be eventually be awarded grades using criterion referencing, but that process was not within the scope of this project.

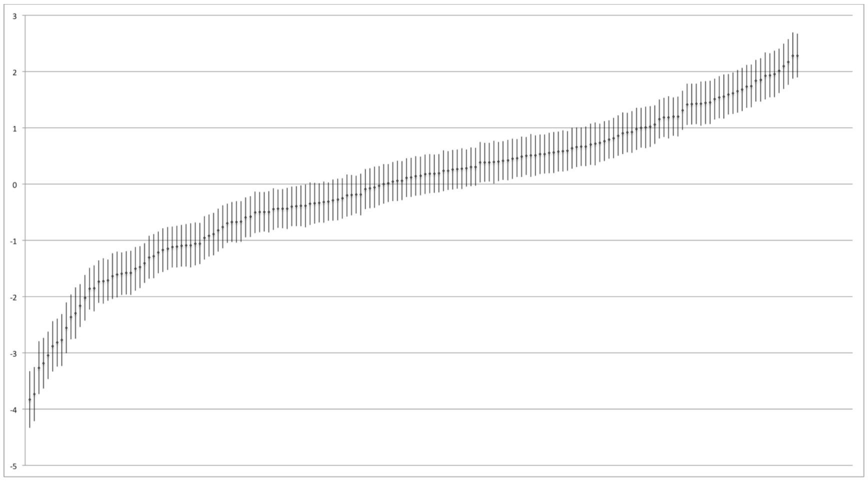


Fig. 17.3 Scaled rank order of student scripts. The horizontal axis shows the 168 scripts from "worst" to "best". The vertical axis is the scripts' parameter values in logits. The standard error of each parameter is also shown.

17.3 Evaluation

We intended to use the students' peer assessment for summative purposes and it was therefore necessary to thoroughly evaluate the process. We undertook a statistical analysis in order to evaluate the consistency and reliability of the rank order of scripts. We also interviewed and surveyed students – and other participants as introduced below – in order to establish which features of scripts influenced them when undertaking pairwise comparisons.

17.3.1 Statistical analysis

To evaluate the students' performance we correlated the rank order they produced with rank orders of the scripts produced by two further groups of participants. One of the groups comprised nine experts (mathematics PhD students) and the other comprised nine novices (social science PhD students with no mathematics qualifications beyond GCSE or equivalent).

The expert group provided a benchmark against which to compare the students' performance. It was expected the expert and student rank orders would correlate very strongly. The novice group provided a control. The participants in the novice group had never studied any advanced mathematics and would thus not be able to use mathematical understanding when making judgements. It was therefore expected the expert and novice groups would correlate weakly at best.

The participants were paid for their time and the procedure was the same for both the expert and novice groups, except for a preparatory activity. The experts were sent the written test and asked to become familiar with it by completing it themselves. The novices, presumably unable to complete the test, were instead sent three student written responses. The novices were asked to inspect the three responses and rank them, as well as they were able, in terms of the students' conceptual understanding of the test question.

Each group then attended a training session lasting 30 minutes. During the training sessions a researcher explained the rationale and theory of ACJ, and demonstrated the "e-scape" website. Two expert participants were unable to attend the workshop and received individualised training instead. The participants practised judging scripts online. Once familiar with the website they were each allocated 94 judgements to be completed within ten days of the training.

Once the judging was complete, the judgements for each group were fitted to a Rasch model. The internal consistency was acceptably high for both the expert group (.97) and the novice group (.99).

17.3.2 Analysis and results of statistical analysis

Spearman's rank correlation coefficients were calculated for the three pairs of rank orders. The outcomes are shown in Table 17.1.

	Peer	Novice
Expert	.628	.546
Novice	.666	

Table 17.1 Spearman rank correlation coefficients between the rank orders produced by the students and the two groups of participants. All correlations are significant at $p < .001$.

The expert and peer rank orders correlated significantly more strongly then the expert and novice rank orders, $Z = 1.670$, $p = .048$. This suggests that the experts and peers were more in agreement with one another about what constitutes a good answer to the question than were the experts and novices. Nevertheless, the significance was marginal and we had anticipated a much more marked difference. We also expected the novice group to correlate much more weakly than it did with either the peer group or the expert group. The relatively strong correlation between the novice and two other groups leads to the counter-intuitive and unexpected conclusion that novices lacking knowledge of advanced mathematics can, to some extent at least, assess understanding of advanced mathematics. Furthermore, it is surprising that the peer and novice rank orders correlate more strongly than do the peer and expert rank orders, albeit this difference falls short of significance, $Z = -.7350$, $p = .231$. Reasons for these unanticipated results are considered later in the report.

17.3.3 Survey

Once the judgement week was complete, the students on the course were sent an email inviting them to complete a short online survey about their experience of completing the judgements (they were informed that two randomly-selected students who completed the survey would each win a book token worth £20). Twenty-five students completed the survey. The same survey was also completed by seven of the expert judges and all nine of the novice judges.

The survey instrument comprised nine items which judges rated using a three point nominal scale. The items were derived from the literature into examiner marking and grading processes (e.g. Crisp, 2010) as well as in consideration of contrasts across the students' responses to the written test. The items were worded as generically as possible such that the instrument could be calibrated for use in future ACJ studies using different test questions and, possibly, in different disciplines. For each item judges were asked to consider whether a criterion had a negative, neutral or positive influence on how they made their decisions when judging a pair of written tests. The nine items are shown in Figure 17.4. The instrument also contained an open response section.

17.3.4 Analysis and results of survey

The results from the students' and participants' responses to the nine items are shown in Figure 17.4. There was no difference between the three groups' mean scores, $F(5,35) = .931$, $p = .473$ and so the groups' responses are combined in Figure 17.4.
Item 5, which asked whether use of colour was influential when judging scripts, was intended as a control item and indeed most responses were "no influence". Items 6 and 8 also addressed surface features, although most respondents were negatively influenced by untidiness. The use of written sentences (item 1), formal notation (item 4), diagrams (item 7) and structure (item 9) were all considered largely positive influences. We were slightly surprised by the uniformity of responses to these items, expecting individual differences such as a preference for formal notation over written sentences. The presence of errors (item 3) was negatively influential and evidence of originality and flair were positive (item 2), as might be expected.

Items 2 and 3 are perhaps the only two that novices were unable to use due to their lack of knowledge of advanced mathematics. This means novices were in fact able to recognise other features when making judgements. This may in part explain why their rank order correlated more strongly than expected with that of the students and experts.

An optional open question asked respondents to "state any other features you think may have influenced you when judging pairs of scripts". The responses revealed three influential features not included in the nine items: completeness (e.g. "whether all parts of the question were answered"), factual recall (e.g. "display

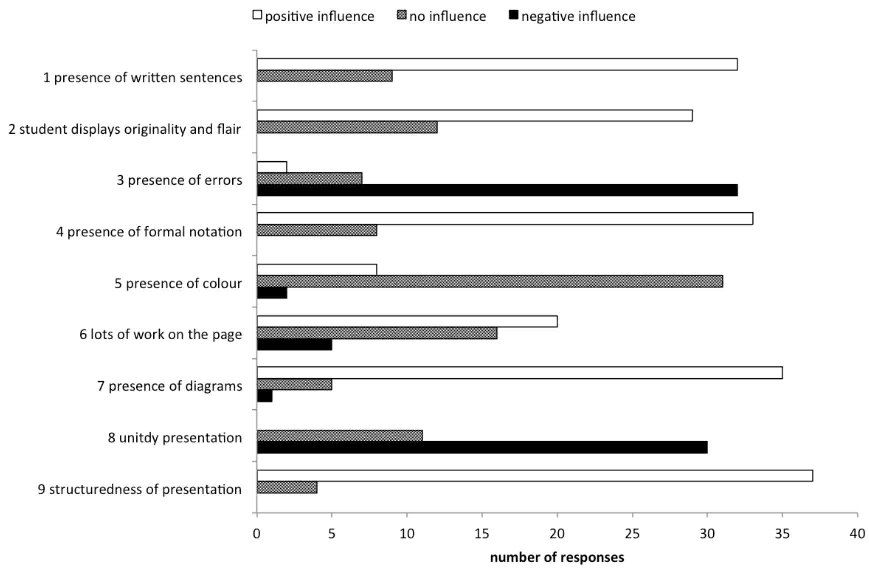


Fig. 17.4 Student and participant responses to the nine items in the online survey.

of knowledge of basic definitions") and vocabulary (e.g. "key words such as flat, smooth, cut had a positive influence"). These items will be included in future adaptations of the instrument.

A second optional open question asked respondents to "comment on your overall experience and feelings about the computer-based part of the conceptual test". Analysis is ongoing but we note here three concerns raised by students. One was that the resolution of the scripts on the screen was too poor to read them properly. Such students presumably did not notice or use the resolution toggle button which overcomes this problem. This feature was demonstrated to the students and highlighted in a support email, and we do not know how many students failed to use it.

Another concern expressed was that not all peers took the activity seriously. One student said, "I do feel that some people may not have to judged the tests accurately as it made no difference to there (*sic*) work. I do understand students should do, however speaking to various students may not have spent the correct time on the computer-based part of the test." This is an astute comment as the quality of the students' judgements had no effect on their final grade. The problem of ensuring undergraduates are properly motivated when assessing one another has been raised in the peer-assessment literature (Topping, 2003), and we return to it later.

Some students commented on the poor quality of some answers, and questioned their peers' ability to assess advanced mathematics. For example, "at least half of the scripts which I read said that the graph was continuous everywhere, when it wasn't. What concerns me is that those people who believe that the graph was continuous everywhere would most probably be marking my own answer wrong." The ability of the students to assess the test can be addressed by statistical analyses, and we discuss further work in this direction below.

17.3.5 Interviews

Semi-structured interviews were conducted with samples from each group of judges. In total nine students, seven experts and three novices were interviewed. Each interview lasted about 20 minutes and was audio recorded and transcribed.

In the interview, the researcher first presented the judge with three pairs of scripts on laminated card. The judge was asked to decide, for each pair of scripts, which was the better in terms of conceptual understanding of the question. They were also asked to give a confidence rating for their decisions on a three-point scale (not at all confident, somewhat confident, very confident). The researcher then asked the participant to talk about each of their decisions in turn using the following three prompt questions:

- How did you decide which test showed the better conceptual understanding?
- Did anything else influence your decision?
- Any other comments about this pair of tests?

Just before the end of each interview the researcher also asked, "How did you find the experience overall?"

17.3.6 Analysis and results of interviews

To analyse the interviewees' judgements of the three pairs of scripts we first identified for each pair which script was the "best" based on an independent expert rank order (see below). This enabled us to designate every judgement made in the interviews as correct (i.e., consistent with the expert rank order) or incorrect. The confidence rating for each correct judgement was scored 1 (not at all confident), 2 (somewhat confident) or 3 (very confident), and conversely each incorrect judgement was scored -1, -2 or -3. We then calculated a weighted score for each interviewee by summing their confidence ratings across the three pairs of scripts. The mean weighted scores across the three groups were 2.14 for the expert group ($N = 7$), -0.44 for the student group ($N = 9$), -0.33 for the novice group ($N = 3$). The experts were the only group to score positively while the students and novices scores were close to zero. This suggests the experts were better able than the peers or novices to judge the scripts, although the small number of participants means we cannot claim statistical significance.

Analysis of responses to the three follow up questions is ongoing and will help us to understand the cognitive processes involved in deciding which of two scripts is the better. Early analysis suggests, perhaps unsurprisingly, that experts, and to an extent peers, focused on mathematical correctness and understanding, whereas novices focused on surface features. To illustrate this, the following responses from each group to scripts A and B are representative:

> *Expert:* First "B" provided more explanation to the answer. "A" just said it is continuous when $x \geq 0$. But "B" said more exactly on the line where $x > 0$ and $y = 0$. And on this line, the function is not continuous and does not have a partial derivative, so I think it confirms "B" is better. And the reason is ok, and also I think "B" said the partial derivative does not exist on the function where it is not continuous.
>
> *Peer:* It was quite hard as they are similar. They have got a lot of the same information on them. The partial derivatives for "A", she said are all 0, and "B" says they don't exist. So I agree with "A". I think they exist.
>
> *Novice:* It was very tight. I am not really confident about this one. But I prefer the way they table the answer in "A" in terms of all elements of the question were approached, they set up the limits, and the bit about continuity, and they got to the partial derivative in a logical order to me. "B" had very nice graphs - although one graph had nothing on. It did not seem as coherent to me.

We note that students' responses to the final question, "How did you find the experience overall?", suggest that they found judging peers' scripts challenging, but beneficial for learning. For example:

> It is hard to judge other people's work . . . Sometimes we as students, we think we understand, but we have to make sure that if someone else reads who has no clue what the concept is, by looking at the question they should be convinced it answers the question. So it is important to write in a good way. It is an improvement for me for my future writing.

17.4 Discussion

In practical terms, the implementation of this novel assessment approach was a success. The scanning and delivery of the scripts to the e-scape system was unproblematic, and no-one in any of the judging groups reported any technical barriers to using the system. All those students who engaged with both parts of the test thus had the opportunity to formulate their own answer to a conceptual question, and to consider the relative merits of responses provided by their peers. Participation was acceptably high - numbers completing both parts of the test were comparable to what would be expected for any other in-class or online test for this amount of credit. In this sense, the goals of the project were successfully achieved.

In theoretical terms, the picture is more mixed. The correlations in Table 17.1, while in the expected direction and statistically significant, are not as anticipated. We expected the correlation between the peer and expert groups to be very strong ($> .9$) and the correlation between the novice and expert groups to be weak ($< .5$). The correlations are also at odds with the experts' superior performance when judging the scripts presented in the interviews, and with their mathematically more sophisticated explanations of how they made their decisions.

The crucial problem appears to have been that the software's adaptive algorithm may not have been optimal when pairing scripts. In other words, a technical glitch meant that the judgements were not informative enough for constructing stable rank orders, no matter how "correct" or internally consistent the judges' decisions. To explore this hypothesis the expert and novice groups are undertaking further judgements. Early analysis suggests the correlation between peers and experts will increase significantly.

Another reason for the relatively low correlation between peers and experts may be due to some students not taking the exercise seriously, or neglecting to adjust the website resolution, or finding the question too difficult to be able to judge the quality of others' answers. We discuss how we intend to address these issues in the next section.

The unexpected results presented us with an immediate practical problem. We had originally intended to use the peers' own judgements for assigning grades. However, the relatively weak correlation between the peer and expert groups caused us to decide not to do this. Instead an independent group of experts, made up of maths and maths education lecturers (and including the course lecturer), has re-judged the scripts and their rank order will be used for grading purposes.

17.5 Further work

Because of the innovative nature of this work, it is currently too early to specify whether and how adaptive comparative judgements will be used as an assessment system in this course or more broadly in the institution. The pressing work required is to test and if necessary improve the adaptive algorithm used to select which pairs of scripts to present to judges. On the basis of previous studies (Jones, Swan and Pollitt, in progress; Kimbell, 2011) we suspect that this alone may go far to improving the peer and expert correlations to acceptable levels. Once improved, we aim to repeat the exercise next academic year.

We will also seek to improve the students' performance by ensuring scripts always load clearly without need to adjust the resolution. Students will also be incentivised to take the exercise seriously by adjusting their grade based on their judging performance. One possibility is to compare their individual judgements with the scripts' positions in a rank order generated by experts. Their performance could then be used to adjust their grade according to agreed levels.

Something that will need to be carefully considered is the question used. On the one hand, the conceptual question used in this instance did challenge the students (many wrote things that were partially correct and partially incorrect), and the responses were very varied so that those students will have seen a wide range of response types. On the other hand, some of the independent experts thought, with hindsight, that this particular question had one problem in particular: the fact that it allowed the students freedom to answer in terms of three different properties meant that it was sometimes difficult to compare two scripts. For instance, how should one compare one script that provides a clear diagram and a correct and well-argued response about the properties of limits and continuity but no information on partial derivatives, with one that has a similar diagram and information about all three properties but contains minor errors? In planning for future tests of a similar nature, we will give more advance consideration to issues of comparability on multiple dimensions.

Addressing these practical issues will also allow us to make further theoretical developments about the use of ACJ for assessing advanced mathematics. We consider the online survey to be the first step in developing a reliable instrument for evaluating the cognitive processes involved in judging. The items will be adapted and extended according to the results and qualitative feedback. We will also increase the rating scale from three to five points to enable more discriminatory responses.

Further ahead we will wish to explore in detail any potential learning benefits that can arise from a pairwise comparisons approach to peer assessment. Many students, and even some PhD maths experts, reported that they felt the exercise was beneficial for learning. A suitable instrument and method will need to be adapted or developed for future studies.

References

Bond, T.G. & Fox, C.M. (2007) *Applying the Rasch model: Fundamental measurement in the human sciences*. Abingdon: Routledge.

Crisp, V. (2010) Towards a model of the judgement processes involved in examination marking. *Oxford Review of Education*. 36(1), 1-21.

Jones, I., Swan, M. and Pollitt, A., (*in progess*) Adaptive Comparative Judgement for assessing sustained mathematical reasoning.

Kimbell, R. (2011) Evolving project e-scape for national assessment. *International Journal of Technology and Design Education*, [online first issue].

Pinto, M. and Tall, D. (2001) Following students' development in a traditional university classroom. In *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education*. Utrecht, The Netherlands: PME, 4, 57-64.

Pollitt, A. (2012) The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*. [online first issue].

Seery, N., Canty, D. and Phelan, P. (2011) The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, [online first issue].

Thurstone, L.L. (1927) A law of comparative judgment. *Psychological Review*. 34(4), 273-286.

Topping, K. (2003) Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.) *Optimising New Modes of Assessment: In Search of Qualities and Standards*. Dordrecht: Kluwer Academic Publishers, 55-87.

Chapter 18
Evaluating Assessment Practices in a Business and Industrial Mathematics Module

Edmund Chadwick and Oana Radu

Abstract *Business and Industrial Mathematics* at the University of Salford is a second year module in the mathematics undergraduate degree. This 20 credit module spans two semesters and the assessment is 100% coursework. A variety of assessments and delivery modes is used. Examples include open-ended problems, problem solving, group work, presentations, report writing, employer seminars and professional studies. The aim of the evaluation study presented here is to investigate the students' perceptions of the various assessments and assessment practices used. We both obtained quantitative measures of the views of the different attributes of the assessments and heard the students' voices in their written comments on the practices they encounter.

18.1 Background and rationale

Business and Industrial Mathematics does not follow the traditional, closed book examination academic route; instead a variety of approaches is used to reflect practices in the workplace for mathematicians. The module aims to prepare students for employment and to highlight the way in which students are expected to use their mathematical knowledge in their future careers and thus the module assesses students for skills related to the workplace (Chadwick, 2011; Chadwick, Sandiford and Percy, 2011). A vital part of this is achieved by exposure to and understanding and experience of work related mathematics.

Throughout the module, seminars from guest speakers on a spectrum of mathematical applications used in industry expand the real world context and give students insights into the world of work and the ways mathematics may be used in their

---

Edmund Chadwick, e-mail: e.a.chadwick@salford.ac.uk
School of Computing Science and Engineering
University of Salford

Oana Radu, e-mail: o.radu@uea.ac.uk
School of Education and Lifelong Learning
University of East Anglia

jobs. These speakers represent a range of different employment sectors and include: DSTL, CMS Intelligent Banking, Manchester Medical Academic Health Sciences (NHS), IBM, an expert court case witness, IMA and the Sellafield OR group. Apart form describing what they do, the guest speakers also describe the group structure of their companies. Talks include topics related to career paths, company profiles and use of mathematics within companies.

These industrial partners identify a number of different attributes which they see as having vital importance, including teamwork, problem solving and professionalism and thus the module attempts to address these attributes in the assessments. There are four different assessments used.

The first is based on teamwork. Students are divided in groups of four, and every member of the group is given a different role: chair, secretary, technical coordinator and task coordinator. Each has responsibilities for particular sections in the final report and the structure attempts to mirror the small team-based approach used by many companies. Two open ended work-related problems are addressed, and each requires the development of a mathematical model. The deliverables include a report detailing the mathematical model, a minutes' book, a project plan and a presentation. The presentation is given to an audience of industrialists, and each team member contributes to it. Following these presentations the audience offers immediate feedback and helps rank the presentations. The academic adviser plays a supportive role and is required only to respond to questions from the group. After the first problem, the team roles are changed to give students greater exposure to different functions. The academic adviser also observes and notes the interaction between members of the groups, how they adapt to their given roles and the change in roles between case studies. Students are given input into the marking, and each student rates the contribution of each team member. From this, individual marks are derived from the group marks. Both the adviser and the students are required to share reflective comments on this experience as part of a wider personal development process for students.

The second assessment focuses on class-based brainstorming to solve a problem. For this part the module leader starts the discussion on the whiteboard by presenting some ideas. The class is encouraged to contribute their own ideas and thoughts. Thus, the whole class contributes to tackling the problem. Each week, a different lecturer leads with his/her ideas and contributions. Each student creates a final report. Thirty percent of this assignment comes from students' in-class contribution to the brainstorming sessions, with the remaining seventy percent from the report.

The third assessment focuses on problem solving abilities. Students are given a bank of games and puzzles. Each week they are able to play these amongst themselves and also take them out on loan. They then focus on particular one of interest to them, and write about the particular game/puzzle in a report. The idea is for the student to be able to describe the position, moves, tactics and strategies for a solution. The description is in the form of a report that represents seventy percent of the total mark. The other thirty percent of the mark comes from their involvement as assessed at the weekly meetings by the module coordinator.

A final assessment is focussed on professionalism. A questionnaire is given to students about the Institute of Mathematics and its Application (IMA) and the process of becoming a chartered mathematician. Another institution compiled this questionnaire on behalf of the IMA. Once students complete the questionnaire, they write a report about the IMA professional body that is assessed by the module coordinator.

18.2 Implementation

The assessments are very varied and broad, and it is unclear exactly how effective each assessment is in measuring students' work-related skills. We sought to evaluate this through a questionnaire to the second year undergraduate students who have taken the module. The questionnaire was designed to give both quantitative and qualitative data, which we evaluate separately below. First, students in both the first and second year were asked to consider which of the following work related attributes are deemed *important and useful* "in a modern undergraduate mathematics degree":

- Professional development
- Mathematical Modelling
- Problem Solving
- Workplace Preparation
- Introduction to Work Practices
- Teamwork
- Employer Engagement

This list was developed in discussion with other academics on the programme and with industrial partners. Students were asked to respond on a five point Likert scale ranging from 'of no importance [or usefulness]' to 'Extremely important [or useful]'. After rating the importance and usefulness of each attribute in a degree, the second year students from the *Business and Industrial Mathematics* module were asked to rate the effectiveness of each of the four assessments in the module at developing these attributes, again on a five point Likert scale.

We also asked the students to complete some open text boxes, giving their opinions on

- Strengths and weaknesses of the module
- Whether they would recommend the module to a friend
- What they would do differently in the module (if anything)
- Whether (and how) they had changed their professional development as a result of the module.

In total 27 students completed the importance/usefulness quantitative questionnaire (16 first years and 11 second years). The 11 second year students evaluated the four different assessments on the module against those attributes and eight of them filled out responses to the qualitative part.

18.3 Evaluation of the quantitative findings of the data analysis

Figure 18.1 shows the mean responses (with standard error bars) of the importance and usefulness of each attribute to a modern degree course. There are no significant differences between the responses to importance and usefulness, so, for the purposes of further analysis, we simply concentrate on importance. There are significant differences between the perceived importance of the attributes $(F(6,182) = 6.52, p < 0.001)$.

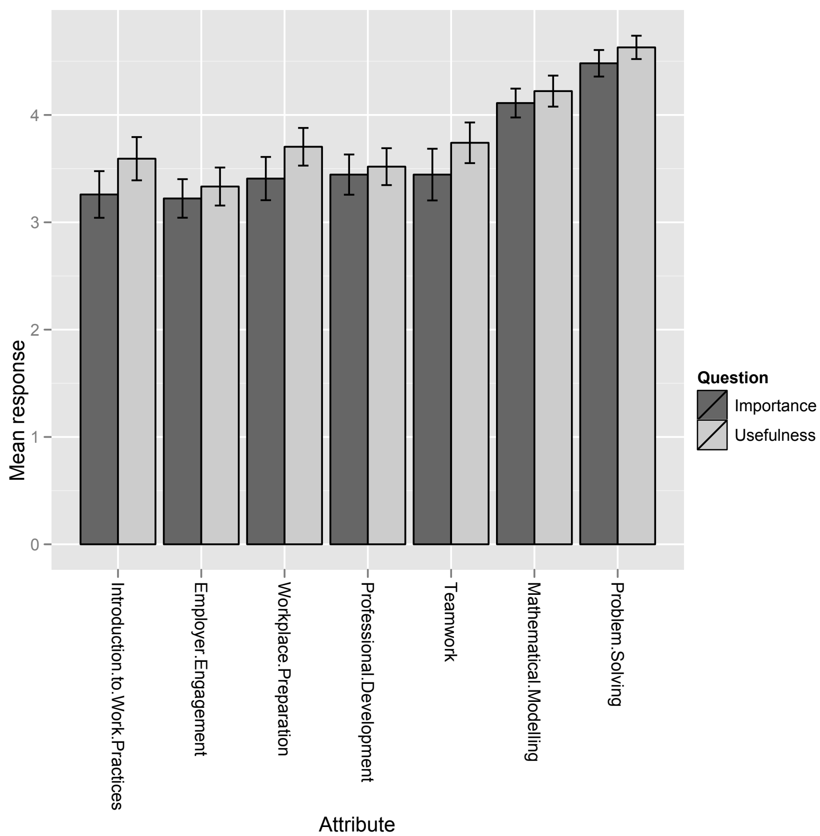


Fig. 18.1 Relative importance and usefulness of the work-related attributes

In particular, it is noticeable that students put the mathematical content (problem solving and mathematical modelling) as more important than professional attributes, such as professional development and introduction to workplace practices. Post-hoc t-tests (with Bonferroni corrections) show that problem solving is significantly more important than all attributes other than modelling and that modelling is significantly more important than employer engagement and introduction to workplace practices (all $ps < 0.05$).

We also found that there were no significant differences between the first year and second year views of these attributes, even though the second years had more experience and had taken the *Business and Industrial Mathematics* module, although the difference between the importance of employer engagement bordered significance ($t(25) = 2.018, p = 0.054$).

Given this pattern of perceived importance of the different attributes, we examined which assessments were seen as more effective at measuring these attributes (Figure 18.2).

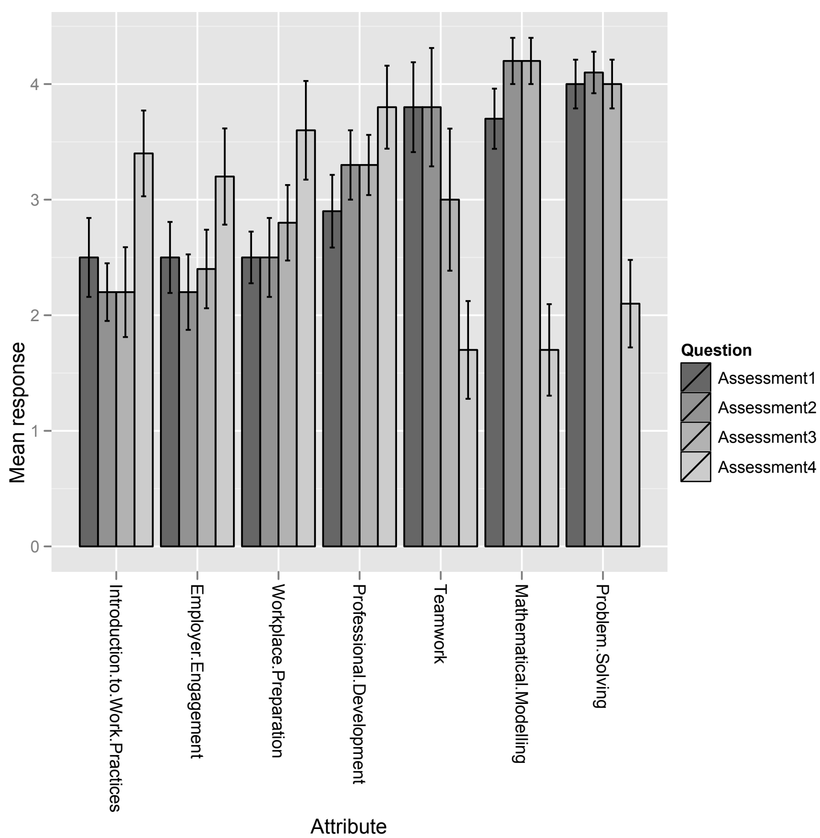


Fig. 18.2 Effectiveness of each assessment over attribute range

A cursory examination of the graph shows that assessment 4 (which focussed on professionalism) appears more effective at assessing the professional attributes of introducing workplace practices, employer engagement and workplace preparation than the other three, though given the sample sizes these differences were not significant. However, assessment 4 was significantly worse at measuring mathematical modelling and problem solving than the other three (all $ps < 0.001$) and worse than assessments 1 and 2 for measuring teamwork (all $p = 0.028$ for both).

There is also the general sense in the graph that students see the first three assessments as having similar patterns of effectiveness across each attribute and that, in general, they are better at assessing problem solving and modelling skills, than the work practice and employer engagement attributes. Again, with small sample sizes, not all of these reach significance, but pairwise t-tests with Bonferonni adjustment show that assessment 1 is seen as significantly more effective at measuring problem solving than work practices, employer engagement or workplace preparation. Assessment 2 is significantly more effective at measuring teamwork, modelling and problem solving than work practices or employer engagement and significantly more effective at measuring modelling and problem solving than workplace preparation. Assessment 3 is significantly more effective at measuring modelling and problem solving than work practices and employer engagement. Finally, assessment 4 is significantly more effective at measuring work practices, employer engagement, workplace preparation and professional development than teamwork or modelling, and significantly more effective at measuring professional development than problem solving (all $ps < 0.05$).

The pattern across all of our data appears to suggest that the skills the students see as more important (problem solving and mathematical modelling) are indeed those which are more effectively measured by three of the four assessments, and, vice versa, the skills seen as least important (work practices and employer engagement) are measured less effectively by those three assessments - though they are covered by the fourth.

18.4 Evaluation of the qualitative findings of the data analysis

The qualitative data analysis presents students' perceptions of the module's strengths and weakness, benefits and views on its assessment strategies. In the following evaluation, the symbol '(n)' after each comment refers to the n^{th} questionnaire sheet.

The data show that students believe that the *Business and Industrial Mathematics* module consolidates and improves their team building skills. It prepares and trains students to further develop their employability skills. The module's configuration encourages students to think about and plan for future mathematics related careers. It helps students to "learn how to work as a team and also how to apply maths skills to real life situations" (3). This is an ideal outcome of the course as students can go on to undertake mathematical as well as non-mathematical related jobs. Students also feel that the module "gets you thinking and perhaps planning for a future career" (7). The assignments' structure appears to help improve students' communication and team skills, such as report writing, keeping minutes and presentation skills. Students believe that the module also improves their mathematics skills. The students state that the course concepts were interesting to work with as these improve students' creativity and perception of mathematics as they learn "to use problem solving and intuition for the earlier assignments" (8). Furthermore, being part of a team improves students' organisational skills. Students report enjoying that this module is not exam based, as they believe such an assessment structure offers more flexibility. They also mentioned the ease with which help and interaction are obtained from the lecturers.

The weaknesses of the module consist of students' failure to understand the relevance of the module for their future career path, the perceived difficulty to work as a team, and their desire to see more research. Students felt that not all the team members pulled their weight for the success of the project in an equal manner and this leads to experiencing "difficulty to work with a group of people who didn't put in as much effort as the others" (5). Ultimately, this creates some difficulties in working together as a team. Students also expressed their desire to have more guidance in organising meetings or in taking meeting minutes. Some students struggled to see the connection between the module and its relevance for future jobs and one even said the module does not present a challenge.

When asked whether or not they would recommend the *Business and Industrial Mathematics* module to a friend, students' perspectives appeared to be divided into four clusters. The first group of students believed the module gave good insights into what mathematics is. It provides interesting insights into mathematics and it is conducive to creating and implementing different solutions paths to math problems because "it was fascinating to see different methods developing to solve the problem" (5). Moreover, the module aims to create links with industry employees. Overall, the module offers an enjoyable mathematical experience and provides a clear and better understanding of mathematics. The second group of students stated that the module was well designed and had fresh perspectives to assessment. The third group of students thought that the module has the ability to open one's horizons as "it also tells you the benefits of joining certain groups" (1). The last cluster of students stated that it helps students in improving their employability skills, such as communication, presentation or report writing.

If given the chance, students would alter the module slightly. They would welcome being given more independence in choosing their groups, would like to see more time invested in creating more enhanced professional projects, and would like to see changes within the assessment structure. Only one respondent believed that the course is perfectly structured.

The students seemed to believe that the course is highly motivational. It helped them in changing their views about the nature of mathematics and even furthered their desire to join the Institute of Mathematics and its Applications. The module also contributed to increasing students' self-confidence in their mathematical abilities and in the team abilities: "now I am more confident in working with other people" (2). It made them consider embarking on graduate studies in mathematics as well as in mathematics related careers. Overall, it made them feel more prepared.

18.5 Discussion, learning and impact

The picture from the quantitative data is, as it should be, of assessments that students perceive as being most effective in assessing precisely those attributes they see as most important (and most useful) in a modern degree course. However, they viewed the assessments quite differently - assessments 1, 2 and 3 all tended to elicit the same pattern of responses (that the more important the attribute, the more effective the measure), but assessment 4 tended to be more effective with the less important attributes.

However, given that the students' views tend to be that the more important attributes are more mathematical (such as problem solving and modelling) and the less important are the direct employment and workplace attributes (introducing workplace practices and employer engagement), it suggests that the balance is about right. It does call into question the extent to which employment related skills may be valued by students even in a module designed to emphasise mathematics in the workplace.

18.6 Further development and sustainability

The qualitative evaluation indicates a direction for future development. Difficulties associated with students' disengagement in working with others in teams need to be addressed. Comments indicate that greater help on how to perform work-related tasks such as writing reports, taking minutes and conducting meetings would be beneficial to students. The problems set to the students could be reconsidered, including increasing the level of difficulty, providing a research focus and further developing the work-related context.

Overall, the module can be deemed successful in that the evaluation demonstrates the assessments were effective in developing the most important and useful attributes in the eyes of the students.

References

Chadwick, E. (2011) Assessing student teams developing mathematical models applied to business and industrial mathematics. Presentation at the CETL-MSOR conference.

Chadwick, E., Sandiford, K. & Percy, D. (2011) Assessing student teams developing mathematical models applied to business and industrial mathematics. *MSOR Connections*. 11(3), 22-24.

Chapter 19
Assessing Proofs in Pure Mathematics

Timothy J. Hetherington

Abstract Many mistakes made by students in coursework and exams arise from poor notation, poor expression of ideas or common misunderstandings. Previous coursework used to assess proof explored their comprehension, clarity of expression, and appreciation of the importance of rigour, but was very time-consuming to mark. Moreover, in the last three years, student numbers have doubled. These issues combined to mean that the assessment used in previous years was no longer viable. This report outlines a project which sought to facilitate the implementation and development of an interesting and innovative assessment on mathematical proof that reduced the marking burden, but that was still educationally rich. The result was a test on mathematical proof which began as a conventional multiple choice quiz, but has now evolved somewhat. This test has dramatically reduced marking time, whilst maintaining student engagement in, and learning from, the process of writing proofs.

19.1 Background and rationale

Since 2008, when I started lecturing, I have been the module leader for a first year module that teaches students about proof. To encourage students to develop their ability to write mathematics I set a piece of coursework (worth 30%) that required students to write a series of short proofs using a number of the standard techniques; direct proof, proof by contradiction, and induction. However, within the framework of each technique the students employed a variety of ideas (with varying success), which meant that each argument, however unorthodox, had to be carefully followed through. For a small group of 40 students this was not an overly onerous task, probably taking between 20 and 25 hours. However, since 2008 student numbers on the mathematics course at Nottingham Trent University have been rising steadily, and by 2011 student numbers were double those of 2008. This increase was the key

Timothy J. Hetherington, e-mail: `timothy.hetherington@ntu.ac.uk`
School of Science and Technology
Nottingham Trent University

driver to rethink the coursework assignment as the huge marking load had made the previous practice unsustainable.

During a Post-Graduate Certificate in Higher Education in 2009, I initially explored ideas on e-assessments, which I thought would naturally address the issue of marking load. These ideas initiated thoughts about how a multiple-choice test on mathematical proof could be developed so that marking the test was easy, but that the test itself was not too trivial. Having built up experience of the common mistakes made by students, it was possible to use these in the development of a multiple-choice test on mathematical proof by basing each incorrect option on a different common error. At the end of the test (worth 10% of the module) students were encouraged to provide feedback on the back of their test papers. There were many positive and supportive comments from the students about this novel mathematical assessment. One student said, “I really enjoyed this new type of test as I’ve never done anything like it before.”

More recently the students sat a second version of the test, which again was well received. The main difference with the second test was that as well as choosing the correct answer, students also had to justify why they had rejected the other three options. This approach was used to eliminate the effect of guessing. After the second test, students were invited to complete a questionnaire and were also given the opportunity to be interviewed about the two tests.

19.2 Implementation

As mentioned, during the previous three years, the coursework assignment required students to write a short series of proofs. In this time I gained experience of common misconceptions and built up a library of false proofs, each of which contained some element of erroneous thinking, be it poor notation, poor expression of ideas, or some more fundamental misunderstanding. For example, students often assume the truth of an equation or inequality that they need to prove. They then work on this equation or inequality, sometimes on both sides at the same time, until they get something they know is true or the same expression on each side of the equals sign. Other popular mistakes include:

- omitting critical information such as what type of numbers are being used;
- misunderstanding concepts or notation such as thinking that ‘divides’ is the same as ‘divided by’;
- placing equals signs at the start of every line of working;
- not identifying the correct assumptions;
- not forming the contrapositive statement correctly;
- proving the converse of what was required;
- not covering all possible cases;
- proving the wrong base case;
- claiming that the induction hypothesis is for all natural numbers.

These errors were then used to create three false proofs for each question, where each false proof highlighted one common error (see figure 19.1 for an example). By basing the false proofs on students' answers from previous years it meant that the incorrect options would seem plausible to some of the students and the popular misconceptions would be highlighted. The provision of formative feedback enables students to learn what is expected of them (Yorke, 2003), and so once detailed feedback on the test was provided, students would be aware not only which option was a logical and well-written proof, but also what common traps to avoid. One aim was to get students to think about how they write mathematics and to appreciate the importance of notation and how they set work out. Many marks are lost in coursework and exams through poor notation, poor expression of ideas, and common misunderstandings. By making students aware in year one of how to write mathematics, it is hoped that improvements may be seen throughout the rest of their course.

1. If $\frac{b}{a} = p$ then $b = ap$ and if $\frac{c}{a} = q$ then $c = aq$. So
$$b + c = ap + aq = a(p+q) \qquad \text{and} \qquad b - c = ap - aq = a(p-q)$$
and it follows that $a|(b \pm c)$.
Incorrect: last bit doesn't follow unless $(p \pm q) \in \mathbb{Z}$.
2. If $a|b$ and $a|c$, then
$$\frac{b}{a} \pm \frac{c}{a} = \frac{b \pm c}{a}.$$
Therefore $a|(b \pm c)$.
Incorrect: is $(b \pm c)/a$ an integer?
3. If $a|b$ then $\exists m \in \mathbb{Z}$ such that $b = am$. Similarly, if $a|c$ then $\exists n \in \mathbb{Z}$ such that $c = an$. So
$$\begin{aligned} b \pm c &= am \pm an \\ &= a(m \pm n). \end{aligned}$$
Since $(m \pm n) \in \mathbb{Z}$, it follows that $a|(b \pm c)$.
Correct.
4. Let
$$\frac{a}{b} = x \qquad \text{and} \qquad \frac{a}{c} = y,$$
where $x, y \in \mathbb{Z}$. Then
$$\frac{a}{x} = b \qquad \text{and} \qquad \frac{a}{y} = c,$$
so
$$b \pm c = \frac{a}{x} \pm \frac{a}{y},$$
which implies that
$$\frac{a}{b \pm c} = x \pm y.$$
Hence $a|(b \pm c)$ since $x \pm y \in \mathbb{Z}$.
Incorrect: $a|b$ is not the same as a/b.

Fig. 19.1 Choices for proofs of the theorem 'if $a|b$ and $a|c$ then $a|(b \pm c)$', with model answers

Three weeks before the test, students were given a list of ten questions on proof. They were told that eight of these would be on their test paper, and that not every paper would be the same. It was hoped that by providing the questions prior to the test, students would spend time trying to write their own proofs, thereby engaging in exactly the same activity that was required for the assignment in previous years. Two examples were provided so that students were aware of how the test would be structured and the task that was required of them. In each example not only was the correct option highlighted, but reasons why the other options were wrong were also given.

The first time that I tried using this test, it was simply multiple-choice; there were eight questions (the easiest eight out of the ten provided), each with four options. Given that the answer for each question was a single letter, copying would be quite easy and so four different versions of the paper were created, each with the same questions, but in a different order. The answer sheet was attached to the back of the paper, again to make it more difficult to copy. Students were explicitly told not to detach their answer sheet as otherwise it would not be known which question paper they had had. With this being an unusual assessment, I was unsure of how long it would take students to complete, and so they were given 80 minutes for the test. However, nearly all had finished in 30 minutes. To mark this test the scripts were first sorted into four groups, each group containing the same version of the test so that each group should have the answers in the same order. It was then very simple to mark, and the whole process of sorting and marking took under two hours for 80 scripts, instead of the 20-25 hours the written coursework had taken. For the first test the mean mark was 67% with a standard deviation of 21%. The distribution of marks was highly skewed; almost half the class achieved a first class mark and three-quarters achieved at least an upper second.

Upon reflection there were some issues with the first test. Given the nature of the topic the options are much longer than one would usually find in a multiple-choice test and so there were not many questions on the paper, only eight in total. This meant that there were only eight marks available, one for each question. Consequently, a student's mark could be affected dramatically by getting one or two questions right or wrong. Given the nature of multiple-choice this means that it was possible that some students had achieved an excellent mark partially due to good luck, whereas other students had achieved a poorer mark due to bad luck.

For example, suppose that a student eliminates two of the options, but cannot decide between the remaining two. If they guess correctly they will achieve 100% for the question, but if they guess incorrectly they will be awarded nothing. However, in each case the student actually has only partial understanding and I wanted to ensure that this partial understanding was reflected in the mark for the multiple-choice question. The wrong answers were not more or less wrong than each other so it was not possible to have different weightings for different answers.

So, for the second trial of the multiple-choice test, to make it fairer it was decided that students should not only identify the correct answer, but they should justify why they had rejected the other three options. Each correct answer was worth one mark, as was each correct justification for rejecting another option. This gives a total of four marks per question and thirty-two marks for the paper. On this version of the test the proportion of marks that could be achieved by guessing has been significantly reduced. Furthermore, comprehension and deeper understanding are tested more rigorously as it is arguably more difficult to identify and articulate precisely what is wrong in an argument than to simply identify the correct argument. The students were given 45 minutes to complete this version of the test. As before, the scripts were sorted into four groups prior to marking, and the process of sorting and marking 80 scripts took about eight hours. For this second test the mean mark was 54% with a standard deviation of 14. The distribution of marks was much closer to a normal distribution, with the majority (58%) of the marks at second class level.

19.3 Evaluation

When evaluating and reflecting upon the delivery, content and assessment of a module it is important to obtain the views of students and to integrate this feedback into the continual cycle of module development (Harvey, 2001). Therefore, to evaluate the project, all students who took the tests were invited to complete a simple online questionnaire with ten questions. The response rate was just over 30% (25/80 students). However, as qualitative feedback is much more useful when trying to make improvements to a module (Harvey, 2001), twelve students were interviewed to establish further their thoughts, ideas, and approaches to the tests. Given that this group of students sat both versions of the test, their views, ideas and insights are highly valuable for evaluating and developing the test further. After all, although the aim was to reduce staff marking time it was not to be at the expense of an educationally rich task; if the assessment is not fit for purpose then the marking time is irrelevant as the most important aspect is the student experience. Since its conception the test has been developed to cater for the learning needs of the students, whilst aiming for a reduction in marking time, which will also lead to quicker feedback. This, in turn, will help to enhance the student experience because for effective learning to take place it is essential to provide quality feedback that is both constructive and timely (Huxham, 2007). Informal feedback after the first test fed into the development of the second test. The more formal feedback presented here, after the second test, will feed into the future development of this assessment.

The first question on the questionnaire asked “Have you ever done a mathematics test like this before?” The response was that 24/25 had never done a mathematics test like this. The results from the other questions are presented in figure 19.2. It can be seen that the vast majority agreed or strongly agreed with the statements presented.

The results from the questionnaire suggest that this was an innovative method of assessment that was enjoyed by the vast majority of the students. Despite the novel assessment method, students felt well-prepared for the test, citing the provision of questions and examples prior to the test as being particularly useful. Students felt that the assessment was a fair measure of their understanding, particularly the second test, and having learnt from the feedback they feel that the whole experience has made them more confident at writing mathematical proofs.

There was also a space for general comments, which most chose not to fill in, but there were some good points raised. Three students said that a detachable answer

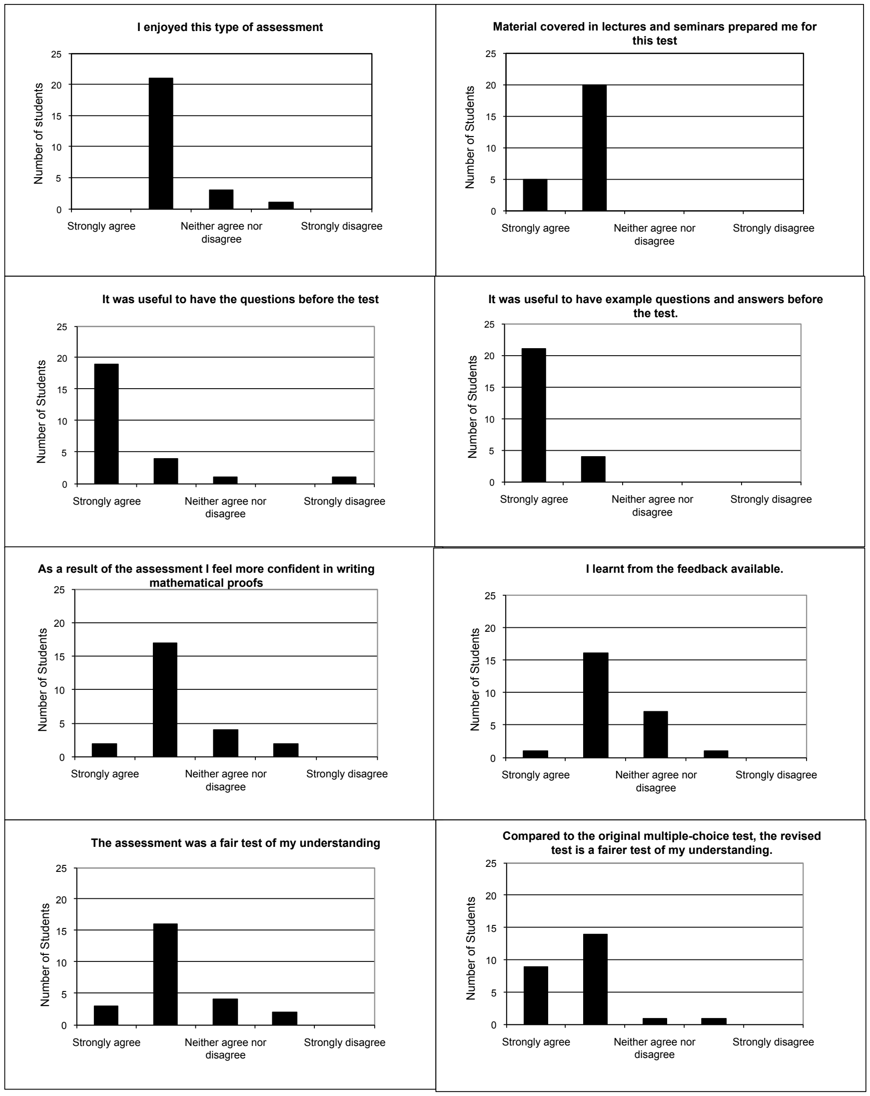


Fig. 19.2 Responses to the questionnaire

sheet would make completing the test easier as currently it is difficult having to flick back and forth to fill in the answers. Two students mentioned that the second test was a better test saying, "Although the first test was easier because one didn't have to explain any reasoning, the second test is a better assessment of pupil's abilities" and "The revised version was better at testing our understanding of proof than the original version." One student commented that they did not think each question should be worth the same mark, another said, "It was quite unclear how we were supposed to write the reason for our choices; were we to write next to every option why it was/wasn't that, or just chose one and justify that one." To pick up on this last comment, it is felt that the rubric on the test paper was perfectly clear; "Stating the correct option will be worth 1 mark AND for each of the three incorrect options a brief justification for why it is incorrect will be worth 1 mark."

The interviews were extremely useful as all of the students who volunteered were quite talkative and had obviously reflected upon their experience.

They were asked to comment on:

- their general thoughts about the test;
- what approach they took;
- whether they would use the same approach again;
- what they thought of the feedback and whether they had learnt from it;
- whether the test raised their awareness of how or how not to write a proof;
- how easy or difficult they found it to put their reasons into words and
- whether they had any suggestions for improving the test.

The general consensus was that the second test, although more difficult, was "fairer than the first one" because "it stopped you from guessing answers" and was "a better way of testing proof". They thought that "having to say why an answer was wrong was a better idea" because "you need to understand to choose the right answer" and they liked the fact that "if you've got half an idea you can still get some marks".

Many of the students had worked in groups to prepare for the test. All of them essentially said that they had "worked through the questions as if it wasn't multiple-choice". They had looked at examples of proof, some making links with other modules, and had tried "to get an understanding of the structure". Some students had tried to think of mistakes that they could make by "working out what the answers might be". It was felt that writing out one's own version of the proof was "good practice". One student said that in the test he "used a highlighter to find differences". All students said that they would use the same approach again.

After the first test each student was given their script, which was attached to the question paper. For each incorrect answer the correct option had been written on their answer sheet. After the second test a mark scheme of model answers was provided so that students could identify exactly why certain options were wrong (as seen in Figure 19.1). The advantages of using model answers are that such feedback can be distributed quickly, avoids being overly negative, and requires the student to actively reflect upon how their own work compares to the model answer (Huxham, 2007). All agreed that the feedback provided was good and that they could not think of anything else that would be useful; "you gave everything we needed". Most students said they had looked at all of the available feedback, and learnt from it. The mark scheme was felt to be most helpful because it gives reasons why options are incorrect; "I learnt why rather than 'it is just wrong'. Now I can pinpoint why."

All students agreed that the test had raised their awareness of how to write mathematics by encouraging them to think about what they are writing; "it made me realise that you have to be pretty accurate with maths proofs" and "it makes you look at notation rather than just the working out". "The way you write it is important. It made us think about why it is wrong. Hopefully we won't make the same mistakes."

Students seemed to "feel more confident about writing proofs now", some adding, "especially in induction questions". Interestingly one student said, "I have more of an idea than if you'd tested in a different way". However, the words of one student provide a reminder about the importance of the continual reinforcement of ideas, "some mistakes are quite obvious, but I think I'll forget".

Most students claimed "it was hard to write reasons" for rejecting certain options; "it is in your head but it is difficult to write down". It was felt that "some were harder than others", with the induction questions being easiest in this sense as explanations such as 'wrong base case' were simple to state.

Ideas for further development were: detaching the answer sheet, making the test open book, giving more time (for the second test), splitting each option into steps to help students pinpoint which step is wrong, writing the answers in the booklet at the end of each question, and giving clearer instructions about what response to give for the correct option.

To summarise the interviews, the students thought the second test was better and fairer as it reduced the element of guesswork and justly rewarded different levels of understanding. In preparation for the tests, students practised writing proofs and tried to think critically about what they were writing. The feedback was useful, particularly the reasons given in the mark scheme, as this was the aspect of the test that students found most difficult.

19.4 Discussion, learning and impact

The outcome of the project is a novel multiple-choice based assessment on mathematical proof that tests comprehension, the ability to identify assumptions in an argument, and raises the appreciation of the importance of rigour. Students felt that this innovative assessment was a fair test of their understanding, but remained a challenging assignment on a topic that many students find difficult. The test is quick and easy to mark, but has maintained student engagement in, and learning from, the process of writing proofs. Moreover, it has raised awareness of common misconceptions and mistakes in mathematical writing. Therefore I feel that the project has not only achieved the intended outcomes, but that after a few modifications (which are discussed later) the assessment will have surpassed my initial expectations.

One of the strengths of the test is that it was based on previous students' work. This means that in the incorrect options the illogical assumptions and mistakes in notation and techniques are exactly the sort of mistakes that students make. Given that the students prepared for the test by writing out their own versions of the proofs, this assignment may have deeper educational merit than simply asking students to write some proofs, as in previous years. In fact, one student commented that "I have more of an idea than if you'd tested in a different way". For this test to work best, the provision of questions and examples prior to the test is critical, which was highlighted in the questionnaires as being particularly useful. The other crucial thing to provide is detailed feedback.

From the questionnaires and interviews it is clear that the students found the test an interesting, enjoyable and rewarding learning experience. The test has helped to raise awareness of the importance of rigour in mathematics, particularly when writing proofs, which should lead to improvements across the whole curriculum. Students mentioned that the whole experience has made them more confident at writing mathematical proofs.

The marks for the (more difficult) second test were encouraging; the mean mark was 54% with a standard deviation of 14. It should be noted that the second test took place three months after the work on proof had been completed. Also, the students suspected that their mark would not count much, if at all, towards their mark in the module, and it was discovered that the test took place on the same day as another test. Therefore, taking these things in account, it is hoped that in future the marks for this test may be a little higher.

Initially it was intended that the test would be an e-assessment so that both marking and feedback would be immediate. However, it has been noted that “e-learning systems are poorly adapted to mathematics” (Smith and Ferguson, 2005: 1). The trouble is that mathematics has its own language, and virtual learning environments are unable to adequately support the necessary mathematical notation and diagrams (ibid.); this was certainly my experience. The University has a virtual learning environment called NOW, which provides access to a variety of tools to enable the creation of e-learning tools. However, when trying to use the e-assessment tool, which can be used to create multiple-choice tests, it was found to be useless for mathematics. Within NOW it is impossible to input any symbols other than those found on the keyboard, and the options for formatting text are very limited. Therefore, in the virtual learning environment, it was not possible to directly type up a well-structured mathematical proof.

However, as it is possible to import pictures into NOW, I created a question and its four options as separate .pdf files, which were then converted into .jpeg and imported. However, this process is cumbersome and time-consuming, and although pictures could be added, it led to issues with pagination and alignment. It was obvious that the technology available was unsuitable for the development of the multiple-choice test on mathematical proof.

One obvious practical issue with the implementation of e-assessment as a summative piece of work for a large class is that more than one computer room would be required to run the test. Also, most people, such as the student who used his highlighter to find differences, would probably prefer to have a paper copy of the test as it is much easier to spot mistakes reading from a printout rather than from the screen.

Now that the assessment has evolved from a simple multiple-choice test it makes implementation through e-assessment impossible: a computer cannot judge whether a reason is correct or not unless it matches pre-programmed permissible phrases.

The development of this novel assessment has been a rewarding experience, and despite the initial cost in terms of time for development and implementation, that time will quickly be recovered in subsequent years due to the huge reduction in marking time. Moreover, the time for implementation and development has been incredibly worthwhile since the product is an excellent assignment that will help with the development of mathematics students for years to come.

19.5 Further development and sustainability

The students mentioned a variety of ideas for improvement and development, many of which coincided with my own. Some of these ideas will be easy to implement, such as slightly changing the rubric to clarify that a correct answer does not require justification. Another simple change is to allow slightly more time for the test. Perhaps somewhere between 50 to 60 minutes would be sufficient, which would mean that the test would still fit into a single lecture slot. It will also be easy to number each line to help students to pinpoint where illogical statements occur, although obviously a valid reason will still need to be given.

One student commented that they did not think each question should be worth the same mark and another suggested writing the answers in the booklet at the end of each question. However, for ease of marking it is felt that neither of these ideas is suitable for implementation.

Another easy improvement would be to place an identifier on each answer sheet to allow them to be detached from the question paper. This will also make the process of sorting the papers into groups easier. However, this sorting process will be unnecessary after the implementation of the following idea.

Students highlighted the mark scheme as invaluable feedback as it included reasons why certain options were incorrect. To ensure that all students engage with this feedback the proposal is to make the assessment into self-assessment, with the answers and marking process forming a discussion within seminars. Not only will this reduce the marking time to a few hours, but it will engage all of the students in a period of reflection by exploring in detail why certain answers were incorrect. It will give students the opportunity to clarify their thoughts, and staff the opportunity to highlight further the popular misconceptions.

In summary, the test will be developed by:

- adding clarification to the rubric;
- numbering each line to aid precise reasoning;
- allowing slightly more time;
- adding an identifier to each answer sheet to allow detachment;
- using self-assessment and discussion within seminars.

As e-learning tools can enhance the learning for those students for which e-learning is a positive experience (O'Regan, 2003), when e-learning systems are available that are well-adapted to mathematics, it would be beneficial to develop multiple-choice tests that concentrate on writing mathematics, but through different subject material. These tests (with no reasoning required) would be well-suited for inclusion in a series of formative e-assessments; formative rather than summative so that the issues with the original multiple-choice test on proof, such as fairness and to some extent the effect of guessing, do not apply. They would be tools for the students to use to guide and inform their own learning and development by reminding and encouraging them to think critically about what they are writing so that they do not slip back into bad habits and are not given the opportunity to forget.

Acknowledgements

I wish to thank Oana Radu for her help with marking the second test and with transcribing some of the student interviews.

References

Harvey, L. (2001) *Student Feedback: a report to the Higher Education Funding Council for England*. Accessed at `www0.bcu.ac.uk/crq/publications/studentfeedback.pdf`

Huxham, M. (2007) Fast and effective feedback: are model answers the answer? *Assessment & Evaluation in Higher Education*. 32(6), 601-611.

O'Regan, K. (2003) Emotion and e-learning. *Journal for Asynchronous Learning Networks*. 7(3), 78-92.

Smith, G. G. & Ferguson, D. (2005) Student attrition in mathematics e-learning. *Australasian Journal of Educational Technology*. 21(3), 323-334.

Yorke, M. (2003) Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice. *Higher Education*. 45, 477-501.

Chapter 20
Towards an Efficient Approach for Examining Employability Skills

Stephen J. Garrett

Abstract A student's approach to an open-ended problem, one with no necessarily right or wrong answer, is crucial to their employability. Indeed the value of a mathematics graduate to an employer is in his/her problem solving skills, and exercises assessing these usually form part of graduate assessment days. The scope for open-ended problems within a mathematics degree is large, yet traditionally these and other transferable skills are not extensively assessed until final year projects. The post-2012 funding shift has significant implications for student recruitment and a fundamental change in the treatment of employability skills is needed in response. The assessment of transferable skills is always possible within extended pieces of coursework, but the marking of these requires a substantial time commitment from staff when class sizes are large. This small study looks at whether it is possible to assess the skills associated with open-ended problems within traditional and time-efficient examinations.

20.1 Background and rationale

With the recent shift in undergraduate funding, the future success or failure of a university department has been placed in the hands of student recruitment and league table performance. The employability of graduates has therefore never been more important. But how should mathematics departments encourage and assess these skills in their students in a time efficient and scalable manner?

Along with communication skills, a student's approach to an open-ended problem, one with no necessarily right or wrong answer, is crucial to their employability. Indeed, the value of a mathematics graduate to an employer is in his/her problem-solving skills and the ability to communicate results, and these are usually the focus of graduate assessment days. There has always been scope for the use of open-ended problems in mathematics degrees, particularly within applied mathematics streams,

Stephen J. Garrett, e-mail: sjg50@leicester.ac.uk
Department of Mathematics
University of Leicester

yet traditionally these skills are not extensively examined until final year projects. A fundamental shift in the treatment of these and other employability skills is needed.

The assessment of problem solving skills is clearly possible by extended pieces of coursework or 'mini-projects' within modules. However, marking these requires a significant time commitment from staff when class sizes are large; as well as being educationally sound, assessment must be practicable. This leads us to question whether it is possible to assess some employability skills within traditional examinations.

This report presents a comparative study of two cohorts of students and their solutions to open-ended problems, both in terms of their approach to the problem and the reporting of their results. One cohort is examined through the use of a 3,000 word mini-project over the semester; the other is examined within a traditional unseen examination. Despite being limited in its scope, this study should be considered as part of an on-going investigation into the effective and efficient assessment of employability skills.

20.2 Implementation

The study compares the summative attainment of two cohorts of students, both enrolled on a module called *Theory of Interest*. The module is an MSc level course within Leicester's MSc *Actuarial Sciences* and MSc *Financial Mathematics & Computation programmes*, however undergraduates are allowed to take it in their final year. Some 37 MSc students (henceforth referred to as PGs) and 87 undergraduates (UGs) took the module this year; these formed the two separate cohorts.

The module is taught in a non-traditional way in that a specifically written textbook is given to the students at the start of the course. Each week a lecture is given on the salient ideas and concepts of a particular chapter, and is delivered with the assumption that the students have already independently studied the material in detail. Furthermore, a weekly problem class is given where questions of examination standard are discussed; again it is assumed that the students have already attempted these questions in advance. The lectures and problem classes are common to both cohorts (PGs and UGs). This approach has the specific aim of instilling independent-learning skills into the students in preparation for professional studies after graduation. This approach has been taken for the last three years and has proved successful and enjoyable, particularly for motivated students. However, a discussion of this aspect of the module is not the aim of this report.

With regards to the assessment of the module, the PGs were assessed by an extended mini-project over the semester and a two-hour unseen written examination consisting of four compulsory technical questions at the end of the semester. The UGs were examined by a three-hour unseen written examination alone, consisting of the same four technical questions and an additional open-ended question, at the end of the semester. The additional examination question was closely related to the PGs' mini-project and both are included in the appendix for reference. Note that the use of compulsory examination questions deviates from typical practice in mathematics programmes and is again intended to prepare students for professional examinations after graduation.

Marking schemes for the open-ended question and mini-project were written to reflect the skills being assessed in each cohort and these are summarised in Table 20.1. Given the very different context of each assessment, the skills do not overlap entirely. The skills common to both cohorts were:

- interpretation of data and information presented in an unfamiliar way (element 1);
- selecting appropriate tools to make an analysis (element 2);
- accurate use of mathematics (element 3);
- forming a decision based on the results of the analysis (element 5).

Both assessment types required the communication of a justification of the methods used in their analysis and justification of their recommendation (element 4). However the target (small business owner vs. academic examiner), style and extent of the required communication were different between the cohorts, and element 4 was not considered as common. The skills being assessed are further discussed in the Evaluation section below.

Element	Weighting	UG cohort	PG cohort
1.	10%	Correct interpretation of information provided	Correct interpretation of information provided
2.	25%	Correct choice of mathematical tools	Correct choice of mathematical tools
3.	25%	Accurate use of mathematical tools	Accurate use of mathematical tools
4.	30%	Justification of approach and recommendations	Justification of approach and recommendations. Appropriate targeted communication style throughout
5.	10%	Clear statement of recommendation	Clear statement of recommendation

Table 20.1 Assessment elements for each cohort

To facilitate marking, detailed mathematical solutions were produced for all calculations students could have attempted, although it was by no means required that students perform all these calculations if adequate justification was given. Given the significant amount of subjective assessment arising from this aspect and the communication skills required in general, the same examiner was used for both cohorts to ensure consistency. It is important to state that the examiner has significant experience within the financial industry and understands the communication skills required, which are very different for those required for the publication of academic papers, for example.

20.3 Evaluation

20.3.1 What skills were assessed in each cohort?

In order to clarify its approach to the development of employability skills at the programme level, Leicester's mathematics department has taken the Great Eight Competencies (Bartram, 2005) as the definitive statement of those fundamental characteristics that underpin job performance. It is useful at this stage to consider how the skills being assessed in this study link back to these competencies. Reference back to this set of competencies is common practice within the department, particularly in the development of new modules with significant skills content.

The competencies relevant to this study are summarised in Table 20.2, where the assessment element refers to that in Table 20.1.

Competency	**UG assessment**	**PG assessment**	**Assessment element**
Leading and deciding	✓	✓	1, 5
Support and cooperating			
Interacting and presenting			
Analysing and interpreting	✓	✓	2, 3
Creating and conceptualising			
Organising and executing	✓	✓	4
Enterprising and performing			

Table 20.2 Bartram's Great Eight Competencies

We consider each of those competencies represented in this paper in more detail:

Leading and deciding: Takes control and exercises leadership. Initiates action, gives direction, and takes responsibility.

Despite the level of communication required, intended audience, and time spent conducting the analysis being very different, both cohorts are required to make and justify a decision based on their analyses. Furthermore, both tasks require students to interpret the information and data given to them in an unfamiliar way: as mathematics students, they are not often confronted with blocks of text containing both relevant and irrelevant information. This competency is therefore associated with assessment elements 1 and 5 and can be compared directly across the cohorts.

Analysing and interpreting: Shows evidence of clear analytical thinking. Gets to the heart of complex problems and issues. Applies own expertise effectively. Quickly takes on new technology. Communicates well in writing.

This competency is relevant to both cohorts and forms the focus of this assessment: selecting and correctly using technical expertise in an unfamiliar situation. It is therefore possible to compare this competency across the cohorts. This competency is associated with assessment elements 1, 2 and 3. Note that the distinction in the communication aspects has been placed under the Organising and Executing competency below.

Organising and executing: Plans ahead and works in a systematic and organised way. Follows directions and procedures. Focuses on customer satisfaction and delivers a quality service or product to the agreed standards.

Clearly the need for time management and planning arise in both cohorts, although with distinctly different time frames. Furthermore, for the UG cohort, the time management issue is for the entire 3-hour examination and the skill within the context of the particular question cannot be isolated for comparison. More importantly, with regards to the customer satisfaction aspects of this competency, the PG task requires the students to write a report with a particular audience in mind. In this particular assessment it is the small business owner, but it could also be a manager, academic, government representative, for example, each of which require different uses of language and descriptions of technical concepts. For the UG task, a simple recommendation and justification is required to be communicated to the academic examiner. The communication aspects of both tasks are therefore fundamentally different and it is not possible to directly compare achievement in this competency across both cohorts, despite it being present in both. This competency is associated with assessment element 4.

20.4 How did the students perform?

A comparison of the summative attainment of each cohort is not an adequate way of determining the success of the assessment for assessing particular skills. However, this information is given in Table 20.3 for completeness.

Assessment element	UG cohort (87 students)	PG cohort (37 students)
1.	61%	49%
2.	58%	65%
3.	68%	72%
4.	65%	53%
5.	85%	75%
Overall	**66% (sd 30%)**	**63% (sd 15%)**

Table 20.3 Summative attainment by assessment element

A comparison of the data in Table 20.3 is difficult because the two cohorts reflect differently qualified students, i.e., the intake for the MSc necessarily filters those students of 2:1 standard and above. This has implications for attainment in the mathematical aspects (element 3) as one might expect the UGs to have weaker mathematical skills than the PGs. Furthermore, the PG cohort has a significantly larger number of overseas students without English as a first language; this has implications for the communication aspects (elements 1, 4 and 5). Both of these issues are anecdotally represented in the data, although no statistical significance is sought for this claim.

Despite having similar averages, the attainment of the UG cohort has a much greater standard deviation around the average then the PG cohort.

20.5 How successful has the UG assessment been in assessing the competencies, relative to the PG assessment?

Clearly the mini-project approach is a better simulation of work: students have access to reference material, have longer timescales, are able to discuss the issues with colleagues, and are required to report the findings with a particular target audience in mind. Furthermore, as the project is completed over an extended period of time, the students are confronted with conflicting deadlines and are required to plan their time and priorities effectively. Such projects should therefore be widely used within mathematics programmes. However, marking the reports is a significant time burden on academics and it is not practicable to include these in every module. In terms of time spent marking during this study, each mini-project took in excess of 30 minutes to read and grade. In contrast, each examination question took around 10 minutes to read and grade.

We proceed to discuss each assessment element from Table 20.1 in turn, paying particular attention to the success of the UG assessment method compared to the PG assessment method. Recall that the aim of this study is to see in what instances the examination-question approach is a suitable alternative to the mini-project approach.

Element 1: correct interpretation of information provided

The same misinterpretations of the information provided in the question occurred in both cohorts, but surprisingly this occurred more often in the PG cohort. Whether this is due to the particular students involved, or was due to the students having more time to overanalyse and confuse themselves over the information is unclear. However, given that the mistakes were common to both, we conclude the UG assessment to be an equally valid means of assessing this competency.

Element 2: correct choice of mathematical tools

The assessment element was twofold. First, the students were expected to think about the investment opportunities available to them and judge their feasibility before performing the mathematical analyses. Indeed a number of the opportunities could be discounted immediately without a full analysis. Second, for those opportunities worthy of mathematical analysis, were appropriate techniques used?

On the whole the PG cohort was less likely to blindly proceed down the mathematical route without reference to practical or commonsense considerations. However, the stronger students from the UG cohort did justify their reasons for not performing analyses better than the PG cohort. Overall both cohorts used the appropriate range of mathematical tools for each analysis.

As with the element 1, given that both cohorts demonstrated a similar range of answers (both correct and incorrect), we conclude the UG assessment to be an equally valid means of assessing this competency.

Element 3: accurate use of mathematical tools

Despite the slightly better achievement of the PG cohort on average with regards to this assessment element, the accurate use of mathematics is clearly assessable in both assessment types.

Element 4: justification of approach and of recommendations (and appropriate communication style throughout)

As discussed previously, it is inappropriate to compare this element across the cohorts. However, in terms of student performance within each cohort, the UG cohort typically performed well, but with a broad range of quality as would be expected from a large UG class. The PG cohort performed poorly compared to expectations, however this was to do with their ability to convey mathematics to a non-specialist audience. We conclude that both assessment types were adequate to assess their very different objectives, with the PG assessment having significant benefits for the assessment of communication skills relevant to employability.

Element 5: clear statement of the recommendation

Implicit in each assessment was the need to summarise the findings. Typically both cohorts performed this well, although different types of summaries were required given the different contexts. Many UG students were in the habit of summarising the main result of their solution as part of good examination technique, and we conclude that the UG assessment is an equally valid means of assessing this element.

20.6 Discussion, learning and impact

This study has been successful in determining that a number of employability competencies can be assessed efficiently within traditional unseen examinations by the use of open-ended questions. This method has the advantage of being a much more time efficient means of assessment compared to the use of mini-projects. Within the confines of this particular study, open-ended examination questions have been shown not to be a valid substitution for the assessment of report writing skills, but that is not to say unseen examinations could not be used to assess other written communication skills. Furthermore, in the context of studying the assessment of all employability skills (as summarised by the Great Eight Competencies), this study is limited. It should be emphasised that this particular study did not cover all the eight competencies and Table 20.2 shows those not considered here. Those missing are typically focused on inter-personal interactions and such skills are best assessed in terms of group work.

Looking at the assessment of employability skills at the programme level, we have shown that assessment of some competencies need not be confined to modules explicitly focused on transferable skills or extended projects, as has tended to be the traditional model. Instead some employability skills can be distributed evenly over a programme and assessed implicitly in the examinations of traditional, technical lecture courses. Explicit modules or elements of modules are required to facilitate the assessment of report writing and group work skills, and the time cost of this is unavoidable.

It is important to note that this study has been concerned with the assessment of employability skills, not the development of employability skills. There is huge value in the provision of problem-based modules to help develop particular skills in students, but it is important that these skills are not left to these modules alone. Regular reinforcement through problem classes and assessment within the examinations of as many modules as possible are to be desired.

A significant barrier to the successful inclusion of employability skills within mathematics and other science programmes is the competence of academic staff. It is clear that academia is only one career option open to graduates, and the skills necessary for success in academia are often distinct from those required in other professions. For example, an academic paper is written entirely differently to, say, a report to a client uneducated in the subject of the report. Despite being skilled in what they do, the relatively narrow skill set of many academics with regards to general employability skills necessitates closer links with employers.

20.7 Further development and sustainability

The results of this study have been fed into the design of a new programme at Leicester, BSc Mathematics and Actuarial Science, to be launched in October 2012. The programme is different from standard mathematics programmes in that it has a clear vocational aim, despite also being academically rigorous. The design of the programme includes the provision for regular and explicit skills modules, and also many modules have been developed with time allocated to the practice of transferable skills, including group work, report writing, presentation and problem solving. The assessment of each individual module will encourage employability skills through continuous assessment marks and also the use of open-ended examination questions as discussed in this paper. The launch of this programme leads to opportunities for further comparative studies in the effectiveness of assessment schemes for employability skills in the coming years.

References

Bartram, D. (2005) The Great Eight Competencies: A Criterion-Centric Approach to Validation, *Journal of Applied Psychology*. 90(6), 1185-1203.

Appendix

Scenario

On 1 April 2011 you are appointed financial adviser to the owner of a small shoemaking business. The business is based in a small village and employs the vast majority of the inhabitants of the village. The village is in a reasonably remote part of Scotland with poor transport links. The business currently makes £200,000 pa profit (after salaries and all operating costs), which is projected to continue for the foreseeable future. Your client's risk-averse nature means that he has a friendly relationship with his bank that provides a business account which earns 2% pa on any deposits, and a rolling loan agreement which charges 4% pa on any borrowings. In previous years the owner simply invested any profits in the deposit account which had a balance of £4.5 million just prior to reinvesting all this in new premises for the factory. The new premises are now fully operational and the business has zero borrowings and cash holdings.

Your client has decided that the company's future profits ought to be put to better use and has brought to you the following investment/business opportunities to advise on:

(a) Immediately invest £300,000 in a 10-year government bond which promises coupon payments of 2% pa.
(b) Immediately invest £200,000 in a 25-year bond issued by a new mining firm which prospects for a rare mineral in a remote part of the Highlands of Scotland. The bond promises to pay coupons of 4% pa.

(c) Temporarily diversify the shoe-making business into high-tech electronics. The venture requires a single initial investment of £200,000 for new machinery, and is expected to break even for the first 5 years before making an annual profit of £10,000, increasing by £10,000 pa for the following 9 years (i.e. 10 profitable years in total). After 15 years the machinery will be obsolete and have an estimated scrap value of £5,000. New employees with specialist skills would be required from the outset and the salary cost of these has been factored into the given data.
(d) Diversify the product range into leather boots. This requires a single initial investment of £100,000 for new machinery and training of existing staff. The new line is expected to break even for the first 2 years before making an annual profit of £10,000 which is expected to increase at 3% pa for the foreseeable future. The machinery is expected to last for many years with only minimal maintenance costs, which are factored into the projected figures.
(e) Invest any profits earned over the year in a diversified portfolio of FTSE 100 shares.

Undergraduate examination question:

Determine which, if any, opportunity your client should invest in. Present any calculations and/or discussions to justify your decision.
Hint: Note that there is not necessarily a right or wrong answer for this question; marks are awarded for sensible discussions and relevant calculations.
Total: 20 marks

Postgraduate mini-project brief:

Write a report to the owner of the business, detailing your advice and recommendations. Any mathematical analyses should be attached as an appendix to the report. Despite running a successful business, you should assume that your client is not financially sophisticated and has only a vague understanding of investment jargon.
Hint: Note that there is not necessarily a right or wrong answer, and marks are awarded for sensible comments and relevant calculations. Consideration should be given to strategic fit, feasibility and impact on the local community, as well as financial issues.
Total: 100 marks

Chapter 21
Mathematics Lecturers' Practice and Perception of Computer-Aided Assessment

Carol L. Robinson, Paul Hernandez-Martinez and Stephen Broughton

Abstract This case study investigates, from the practitioners' point of view, the advantages and disadvantages of Computer-Assisted Assessment (CAA), and how lecturers that use this type of assessment deal with the issues involved. Data were collected through a questionnaire and follow-up interviews with lecturers that use CAA in their first year mathematics modules at a large university. Some of the advantages lecturers mentioned were saving time in designing and marking tests and giving feedback to large groups of students, student motivation, socialisation of learning and peer support, and students having a more relaxed way of being assessed (when tests were not invigilated). On the other hand, lecturers noted the procedural nature of CAA tests, poor quality feedback and the inability to change an "antiquated" system. Lecturers using CAA make compromises in order to retain the advantages of the system by, for example, reducing the contribution that CAA tests have in the overall assessment scheme or testing conceptual understanding through other means; but it is clear that they would welcome a simpler, more effective system that could address the shortfalls of the current one.

21.1 Rationale and aims

Mathematics lecturers at the target university are in the position of being able to utilise Computer-Assisted Assessment (CAA) without the need to develop their own questions. Two projects undertaken a few years ago by colleagues at the university and elsewhere have resulted in question banks containing thousands of questions ready to use (HELM, 2006). This project evaluates the issues arising for lecturers who use these resources as a method of assessment.

It would appear at first sight that the ready availability of CAA questions is an extremely efficient way of assessing hundreds of first year students and would be wel-

Carol L. Robinson, e-mail: c.l.robinson@lboro.ac.uk
Paul Hernandez-Martinez, e-mail: p.a.hernandez-martinez@lboro.ac.uk
Stephen Broughton, e-mail: s.broughton@lboro.ac.uk
Mathematics Education Centre
Loughborough University

comed by all involved. Question banks are available for both practice and coursework tests and lecturers are freed from marking students' work. The workload for lecturers is minimal, as dedicated e-learning technicians are available to upload tests and the computer software provides pre-prepared feedback to the students and summary statistics for the lecturers.

However, all is not necessarily as straightforward as it might appear. For most large classes it is not possible to invigilate the coursework tests due to the lack of availability of computer labs for this purpose. If students then take the tests in their own time, some lecturers and departments are concerned about plagiarism and, in these cases, paper-based invigilated versions of tests may need to be prepared and marked, thus reducing the efficiency of the system. Other lecturers are concerned about the questions that are available for use. Sometimes they do not fully cover the required syllabus, but the steep learning curve and associated time involved in developing new questions is prohibitive and so lecturers may be tempted to "make do". Other concerns involve the nature of many CAA questions, which seems more suited to testing techniques or procedures than conceptual understanding.

This project addresses the following research questions:

- RQ1 How is CAA implemented in first year mathematics modules for mathematics and engineering students at the target university?
- RQ2 Why are lecturers using CAA?
- RQ3 What are the lecturers' perceptions of issues arising?
- RQ4 How are lecturers dealing with these issues?

21.2 Background

Lecturers in the mathematics and mathematics education departments of this university are responsible for the development and delivery of the teaching of mathematics and statistics for undergraduate mathematics students and most undergraduate engineering students. This study focuses on CAA delivered to first year students. Currently there are over 200 first year mathematics students and approximately 600 first year engineering students taught by staff in the two departments.

The question banks were developed for engineering mathematics modules as part of the HELM project (Harrison, Green, Pidcock and Palipana, 2007). These cover all the first year engineering mathematics topics such as vectors, complex numbers, matrices, differential equations, etc. In parallel to this, staff in the mathematics department developed CAA questions for the two main first year modules for undergraduate mathematics students, namely calculus and linear algebra. The question banks are each separated into two parts – one for practice tests and one for assessed coursework tests. Lecturers may choose to use CAA for both practice and coursework tests or for just one aspect or not to use it at all. Lecturers can also choose whether to provide detailed feedback or simply indicate which questions have been answered correctly/incorrectly. Figure 21.1 shows a typical question and Figure 21.2 the specific feedback provided to students.

Determine the real value of k for which

$$\frac{d^2y}{d^2x} + k\frac{dy}{dx} + 16k = 0$$

has the solution $y = xe^{-4x}$

Enter your answer in the box provided.

This question is worth 2 mark(s)

Fig. 21.1 A typical question on differential equations for first year engineering undergraduates

Constant coefficient ordinary differential equation have solutions of this kind when the auxiliary equation

$$\lambda^2 + k\lambda + 16 = 0$$

has a double root for λ.

This will occur in this case when $k^2 = 64$.

Note: $\lambda = \frac{-k \pm \sqrt{k^2 - 64}}{2}$

So the value of $k = 8$ will lead to the solution $y = xe^{-4x}$
Had we taken $k = -8$ a solution of the form $y = xe^{4x}$
would have been obtained.

Fig. 21.2 Feedback for the differential equations question

21.3 Methods

In order to answer the research questions, the approach adopted was that of a questionnaire followed up with interviews.

21.3.1 The questionnaire

A questionnaire was developed and then piloted with postgraduate students and one lecturer of a second year module that uses CAA. All thirteen lecturers teaching mathematics modules to first year mathematics and engineering students were then invited to complete it. Four of these were eliminated from this study, since currently they do not use CAA with first year undergraduates.

The first section of the questionnaire (questions 1-7) focused on how each lecturer used CAA in his/her module and covered aspects such as availability of practice tests, format of coursework tests and whether these were paper-based or invigilated, the type of feedback provided and the lecturer's policy on collaboration for tests. Responses from this enable us to address RQ1. Question 8 focussed on the lecturers' perceptions of the type of mathematical understanding that CAA tests. Question 9 explored reasons for using CAA (RQ2) and question 10 focussed on authoring of CAA questions. Question 11 gave lecturers the opportunity to provide additional comments. Finally, respondents were invited to indicate their willingness to take part in a follow-up interview.

21.3.2 The interviews

Those lecturers that indicated willingness to take part in the follow-up interview were invited to suggest a time that would be most suitable to them. Some were conducted in the lecturers' own offices and others were held in rooms booked in the mathematics education department building. The interviews were semi-structured, using the questionnaire as a basis from which to establish lecturers' detailed reasons for their choices.

The interviews lasted between 27 minutes and 54 minutes (median time 34 minutes). The questions addressed in the interviews were:

- Why do you use CAA?
- Why is CAA set up this way in your module? What changes might you make in the near future?
- What does CAA test?
- How have your interactions with students and other lecturers influenced the way you use CAA?
- What are the reasons for your policy on collaboration between students in CAA activities?
- In what ways do students collaborate in CAA exercises?

21.4 Results

We first focus on the results from the questionnaire and then turn our attention to the interview findings.

21.4.1 Questionnaire findings

The use of CAA

All nine lecturers use CAA practice tests with their students and seven of these use CAA for coursework tests. One lecturer uses paper-based tests in order to ask more challenging questions: this lecturer allows students to access the practice tests, since they are preparation for some of the paper test questions. A second lecturer also makes practice tests available to students, but they are not followed by a test. Three of the seven lecturers that use the CAA summative test invigilate the test in a computer lab; and the paper test is invigilated in a lecture theatre. The remaining four lecturers allow the students to take the tests at a location of their choosing and in their own time, within a specified time-period (usually two or three days). Lecturers' perceptions about the need for invigilation were explored in more depth during interviews. Availability of practice tests also varies and details of this and the information regarding invigilation and use of paper-based tests are summarised in Table 21.1.

		CAA test		**Invigilated paper-based test**	**No test**
		Invigilated	Non-invigilated		
Access to practice test granted more than one week before a test	Access to practice tests granted after the test	0	2	1	1
Access to practice test granted up to one week before a test	Access to practice tests granted after the test	0	1	0	0
	Access to practice tests not granted after the test	3	1	0	0
Totals		3	4	1	1

Table 21.1 Summary of implementation of CAA

Reasons for using CAA

In question 9 lecturers were presented with nine possible reasons for using CAA and asked to select from a scale of 'strongly agree' to 'strongly disagree' with the given statements. Not all lecturers responded to all the statements.

I use CAA with students because . . .	**Strongly disagree**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
it is easy to set CAA tests for my students	1	1	0	4	2
it was used by a previous lecturer	0	0	0	4	4
I am encouraged to by the department	0	0	6	1	1
students receive immediate feedback	0	0	2	4	3
students receive good quality feedback	0	2	3	3	1
CAA frees up time	0	0	0	5	4
CAA is convenient	0	0	0	5	3
CAA provides students opportunities to practise	0	0	1	4	4
CAA provides students motivation to practise	0	1	2	5	0

Table 21.2 Why lecturers use CAA – questionnaire responses

From Table 21.2 we see that although there is strong agreement that CAA frees up time, is convenient, provides opportunities and motivation for students to practise and provides immediate feedback, there is disagreement on the quality of feedback received by students. This conflict, between what the system provides for their students in terms of feedback, and what they might wish to provide will be explored in more detail in the analysis of the interviews. There is also disagreement regarding the ease of setting up CAA tests – this will also be explored in the analysis of the interviews.

Setting questions

The current bank of questions provides a permanent source of questions to set tests, but lecturers disagree when asked whether these questions provide students with sufficient challenge (2 believe they do; 5 believe they do not; the remaining 2 neither agree nor disagree). While it would seem that developing new questions would then be worthwhile, some lecturers feel that developing new questions takes too much time (4 agree; 5 neither agree nor disagree) and some feel that developing questions is too difficult (3 agree; 6 neither agree nor disagree). Only three of the lecturers have attempted to write their own CAA questions and fewer still (two) believe it would be worthwhile to learn how to do so.

Thus there is an issue and potential conflict here. Over half the lecturers questioned do not believe the questions provide enough of a challenge. However the alternative, of developing new questions, is not a route they are adopting. We explore this in more detail in the interviews.

Collaboration of students

All of the lecturers are happy with collaboration in the practice tests, but only three explicitly encourage students to collaborate. Those lecturers that have an invigilated test (either online CAA or paper-based) prevent collaboration by enforcing "exam conditions". However, the four lecturers that do not invigilate the CAA coursework tests do not wish students to collaborate, but they do not communicate these wishes to the students.

Testing of recall, procedural ability and knowledge of concepts

In question 8 lecturers were asked to indicate the extent to which they agree that CAA tests recall, procedural ability and knowledge of concepts.

CAA tests my students' ...	**Strongly disagree**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
ability to recall mathematical facts, rules and equations	0	3	1	3	2
ability to carry out mathematical procedures and methods	0	0	1	5	3
deeper understanding of mathematical concepts	0	3	5	1	0

Table 21.3 What lecturers believe CAA tests of students

From Table 21.3 it may be seen that there is strong agreement that CAA tests recall and procedural ability, but not so in the case of knowledge of concepts. Interviews were used to probe this and try to ascertain how much of an issue this was for the lecturers concerned.

Most lecturers provide detailed feedback to students through CAA practice tests (seven; and two do not). Of the lecturers that use the online coursework test, two lecturers provide more detailed feedback at this opportunity as well as in the practice tests.

21.4.2 Interviews findings

Reasons for using CAA

Established practice

All six lecturers that were interviewed say that the reason that led them into this practice was that they inherited it from previous lecturers. One lecturer said, “That’s the way it has been done” before adding, “One of the reasons that I have been supportive of using CAA is that it was in operation here” (P1) . Other lecturers suggest similar histories: “that’s what I inherited” (P3); ‘the honest answer is probably because I’ve inherited it that way” (P4); “I’ve inherited it with the Calculus module that was taught previously” (P6).

Departmental influence

The interviewed lecturers have discussed the use of CAA with other lecturers at some point. However, the extent to which these discussions have had an effect on their teaching and use of CAA differ. One lecturer said, “When it comes down to it, I use it because I’ve been told to use it” (P5). Another lecturer suggests she would not have implemented CAA had it not been for her colleagues’ influence: “I suppose [*discussions with other lecturers have*] been a strong influence, because I hadn’t used it before” (P1). When asked whether discussions with other lecturers have influenced his use of CAA, one lecturer replied, “Not much” (P4).

Saving time

There are two aspects of CAA that help save time. First, CAA handles the distribution and marking of tests and returns feedback with little further human intervention. Second, the task of setting the tests and questions can be shared with others. Most lecturers are wary of spending a disproportionate amount of time on assessment. This is particularly troublesome with large student groups: “If I’m going to consider [*giving*] a written piece of assessment . . . to 200 students, there’s just too much marking involved” (P5). CAA offers the means to distribute and mark students’ work without further input: and this is convenient, “It certainly frees up your time; it’s convenient” (P3).

The university has dedicated staff that work closely with CAA systems and set up access to the tests. For the lecturers, this means that they can simply choose questions from the question bank and ask someone else to compile the test: “what’s particularly good about them . . . is that, more or less, somebody else does all the work” (P1) and “the system is all set up, so I don’t have to do any of that” (P3).

Student responsibility

Some motivation for using CAA is to encourage students to accept some responsibility for the learning they do. Lecturers say: '[*With CAA*] they have to take a strong degree of responsibility for their own learning" (P1); "with the computer tests, I think you encourage them to go and do some work" (P2); "it does give them opportunities to practice" (P3). And another lecturer believes that students seem to relish this responsibility: "the teaching coordinator asked me to have a look at how many students actually do that [*attempt the practice tests*] and it turns out that they do this quite a bit . . . For me it was a surprisingly high average of how many times students do these tests" (P4). For some lecturers, this is part of a wider approach to fostering a mature approach to learning at university: "a practice test is for not just practising getting the right answers, but understanding what the questions are, how to go about them and, if they have a problem, to find out what it is they have a problem with" (P5).

Lecturers' perceptions of the issues

Antiquated system

Some limitations of the system are attributed to its age. Some lecturers commented on this aspect of the CAA system when highlighting particular problems with it: "that we have to produce the question in a 'jpeg' is, I think, rather odd. I don't know where it comes from. Maybe it just shows that the system is ten or fifteen years old" (P4); "the discussions that I've had with lecturers have been along the lines of, 'This is such an antiquated system.' " (P6).

Although age does not necessarily render an assessment technique such as CAA useless, the emergence of other, younger systems attracts attention. "I have been to talks where . . . you get potentially better feedback . . . and that sort of thing is very appealing" (P3), in contrast to "I think there probably are systems which would make it not so onerous, but the one that we currently have is just a nightmare" (P6).

The burden on time

One of the key issues that lecturers face when developing new questions is that it is time-consuming and involves a steep learning curve. One lecturer examined the possibility of developing questions to suit her teaching group, however she found that considerable effort was required: "There is a system where you can write your own questions, but that is a lot of work. I think it's five hours for one question, and you have to really learn the system" (P5).

Many lecturers would like to change the questions but feel unable to. The large bank of questions is a treasured and time-saving resource: "if we use something else, then that means that we've got to leave behind the question bank that we're using" (P6). However, the existing questions no longer provide sufficient challenge for the students: "they don't offer as much of a challenge as I want" (P1); "the level is similar to the ones in the tutorial. And as I said, it's just a matter of repeating" (P2).

Testing students

There is a growing desire to test students more deeply: "I'd be even happier if I could push to more complicated and more conceptual questions" (P4). However, the existing questions do not always provide this, hence the need to develop new questions. There is an emerging acknowledgement that CAA is most effective at testing procedural ability: "it's quite effective at making sure that they can carry out the procedures" (P6). It falls short of being able to test conceptual knowledge and recall: "I don't think it tests their recall, because they can have all of their materials in front of them" (P6).

There are further concerns that Computer-Assisted Assessment conditions the students into learning in a particular way. The formulaic nature of the questions and the CAA testing routine encourages students to "just do enough repetition" (P6) until they are proficient in those questions. There is a danger that this miscommunicates to students the nature of other assessments: "if you, for example, prepare students under certain conditions, when it comes to the final examination, they get used to that" (P2).

This seems evident from the requests that students make to lecturers to have the practice tests available in the leading weeks before the final exam: "I had a lot of requests before the final examinations where students asked me, 'Is it possible to have all these tests, the practice tests, on the machine or on the system?' " (P2).

Obtaining feedback

Although CAA can help students become familiar with the procedures they need to learn, lecturers believe that the feedback they receive is not necessarily helpful. For the most able, CAA confirms to students that they have carried out the procedure correctly: "I am sure there are students who think it's [*CAA is*] very effective because they are getting 85%. They can tick it off. They know that they are doing well" (P5). However, CAA might struggle to provide the feedback necessary to facilitate understanding in weaker students. These students are presented with solutions that look similar to ones they have already experienced in lectures, giving no extra support than the one they already had: "the feedback that we're giving there isn't much more than another worked example, as you find in the lecture notes, or as you find in the textbooks" (P4). Consequently, lecturers tend to be dissatisfied with the feedback as it lacks the ability to respond to students' work: "I don't think it's good quality feedback in the sense of being individual, or being able to give hints"

(P3); “I do say to my students in the lecture that that level of feedback is not what I would like them to have” (P4). Some lecturers are also unhappy about the feedback they receive from the system, which ideally could help them identify what has been understood and which topics need more focus. One lecturer noted, “the CAA tests themselves . . . would not tell me what students can and cannot do,” adding “if they’re getting a low score, I cannot tell what they know and don’t know” (P5). Another agreed: “I have very limited knowledge of what the student has actually done” (P4).

There are also issues with the scores that students receive. The lecturers report that the students feel the scores awarded by CAA do not necessarily reflect the knowledge they have. One lecturer was disappointed that since CAA cannot interpret the intermediate work that students have done, students are harshly penalised if their answer does not match the solution: “They were frequently failing on the detail . . . and getting no marks, even though they were doing the right thing” (P5). Furthermore, there have been instances of mistakes in the CAA questions that only recently have been identified: “Embarrassingly, even in the problems that we actually have run for ten years . . . students did find mistakes” (P4). Some lecturers believe that CAA is seen as the epitome of fairness, since no judgements are made on students’ work: “because we are required to give them a mark, we at least want it to be as consistent and as fair . . . as possible” (P6). However, perhaps due to this perception, some students do not challenge the marks they have been awarded: “Nobody ever complains about the marks on a Computer-Aided Assessment, probably because they’re given by a computer” (P6).

Coping with large group sizes

Many lecturers are keen to maintain CAA as it helps with the assessment of large cohorts of students: “CAA testing didn’t come out of a bad intention. It was a drive to make things work for large groups” (P5). However, the consequence of using CAA with large groups is that there are no computer laboratories large enough to accommodate every student: “it’s 200 people. I’m not even sure we have a computer room that big” (P3). For these groups, invigilating the summative test is not possible. Consequently, students have much more freedom over their environment while performing these tests. This is an advantage for some lecturers, since students face many assessments in their first year: “they sort of like doing things in their own time, on the computer, in their own room, or whatever. I think it’s less stressful in many ways than other forms of assessment. And I think the poor first years are so stressed out most of the time that I think to make other forms of coursework that they have a test in . . . would be much more arduous” (P6). However, without invigilation, lecturers cannot be sure how the test was completed: “how do you know who’s done it? How do you know that they’ve done it on their own? How do you know if they’ve copied from somebody else or from the book?”(P1). For the practice tests, most lecturers encourage collaboration. However, university regulations prohibit collaboration in summative testing. There are concerns that some students continue to collaborate when taking the summative test. Since invigilation is unviable, lecturers are made oblivious to the activities of their students: “To be honest with you, I have no idea” (P2); “I have no idea, I couldn’t tell you’ (P3); and, “I have no idea where the answers come from” (P4).

Dealing with issues of CAA

Changing the place of CAA

With the acknowledgement that CAA helps students to learn procedures and methods, some lecturers have made the practice tests available throughout the year so that students can revisit the material and be tested on the content prior to the module exam: “when students have requested having it [*practice tests*] again for revision purposes, that has been done” (P1); “they’re really meant to be a study aid and a ‘Have you really understood things as well as you think you have?”’ (P6). Some lecturers are able to invigilate the tests, where groups are smaller. In larger groups, where invigilation is not possible, lecturers have tried to minimise the problem by reducing the contribution that CAA tests have towards a student’s module mark rather than looking for ways to invigilate the tests: “I’m not sufficiently worried about it to really make my own life and theirs much more difficult by starting to run it as an invigilated test” (P3). In that way, students remain sufficiently motivated to use the practice tests and the impact of any cheating is made smaller: “For 2.5%, I just don’t think it’s worth putting up a major police operation to find out what students really do”(P4).

Some modules have been adjusted so that conceptual understanding can be assessed in other ways. One lecturer explained how the introduction of a project has also affected the weighting of CAA in the overall module score: “we actually restructured the whole assessment and reduced the number of CAAs and also the weighting, because the coursework then took that weighting away” (P5). The rebalancing of assessments in modules has enabled lecturers to focus on certain aspects of understanding with each assessment tool. While exams can test recall, procedural and conceptual knowledge, CAA can focus on procedures while other coursework can explore deeper understanding: “[*CAA*] is quite effective at making sure that they can carry out the procedures. And since I want them to be able to carry out procedures and I want them to be able to do the conceptual things, I just test the procedures through the computer courseworks, and I test the conceptual things through written courseworks” (P6).

Future practice

Lecturers still have certain misgivings about the CAA system, to the extent to which they describe it “antiquated” (P6), “awful” (P4), “a nightmare” (P6) and “poor quality" (P3), for instance. Other lecturers were less scathing, describing CAA as "good" (P5).

There is a general consensus that further change to practice and assessment structure is not necessary. However, it appears that the focus of change would be of the system, rather than of practice: "I guess it's only a matter of time before we get some other kind of system, which will undoubtedly do some things better" (P3). Such a change would be an upheaval ("I think that the change to the system is not something that can be made at a single course level, because there are other modules that use the system" (P6)), and thus changing the system requires a carefully made decision: "a number of people in the department are looking at alternative software that we might want to use. So far, they haven't come up with a decision" (P4).

21.5 Discussion, Learning and Impact

Our questionnaire and follow-up interviews show a number of issues that lecturers face when using CAA.

Lecturers have to deal with these and decide if the advantages that CAA gives are greater than the disadvantages. Obviously, the fact that a considerable number of lecturers are still using CAA shows that they think CAA is somehow useful in assessing their students' mathematical learning (or some aspect of it). Some of these lecturers have to make compromises, at least until something "better" comes along. In making these compromises, they have sought to minimise the disadvantages by reducing the contribution that CAA tests have in the overall assessment scheme. Also, if CAA only tests procedural understanding, then conceptual understanding can be tested through other means, for example, a project. Some of these lecturers have also made available to students the practice tests throughout the year in the hope that students will use them as a complement to their studies and not only as a vehicle to get a "good" mark in the exam. One thing that was clear from the data was that lecturers are very aware of the downfalls of CAA and, whilst trying to minimise these, they would welcome a simpler, more effective system that takes into account the advantages that the current system gives but also addresses some of the issues presented here.

References

Harrison, M.C., Green, D.R., Pidcock, D. and Palipana, A.S. (2007) HELM Educational Transfer. *MSOR Connections*. 7(1), 20-22.

HELM (2006) *HELM - Helping Engineers Learn Mathematics.* [online] Available at: `http://helm.lboro.ac.uk` [Accessed 16 April 2011]

Chapter 22
Use of Audience Response Devices for Formative and Summative Assessment

Paul Hewson and David Graham

Abstract We conducted focus group and survey evaluations of students' experiences and expectations regarding the use of audience response 'clickers' in statistics and mathematics classes. There was evidence that clickers can enhance learning by encouraging work on problems, by allowing all students time to work on problems, by providing quick feedback on success in problem solving and by allowing the lecturer to adapt the lecture according to common problems. We consider the types of assessment which can encourage engagement within lectures.

22.1 Background and rationale

Personal response (clicker) systems have received a tremendous amount of attention in the learning community – especially in the USA – and their use is widespread at many levels of education. There is some evidence that, if used appropriately, such systems can improve student engagement, understanding and performance (Bode et al., 2009; Robinson and King, 2009). Typically, lectures are broken up by brief sessions of formative assessment using the clickers. This enables the teacher to 'reset' the class clock and can help maintain students' concentration levels throughout the session. The use of clickers in summative assessment is relatively less common and the literature is again not clear on the issues associated with marks being formally associated with clicker tasks (Chin and Brown, 2002; Kay and Sage, 2009). Our experience is that students actually like frequent assessment if it is manageable in quantity and is accompanied by rapid feedback. Clearly, there is potential for clickers to provide this. Also, perhaps surprisingly, we have had students suggest they would prefer to take in-class tests (that is, summative assessment) using these devices. This study investigates the attitudes of mathematics and statistics students to the use of clickers in the classroom environment in general as well as specifically looking at the possibilities of formally assessing the subject using this technology.

Paul Hewson, e-mail: `paul.hewson@plymouth.ac.uk`
David Graham, e-mail: `d.graham@plymouth.ac.uk`
School of Computing and Mathematics
Plymouth University

22.2 Implementation

We ran a focus group with volunteers from a second year applied statistics (Regression Modelling) class where clickers were in current use and then designed a survey which was sent to members of this class, members who took this same module the previous year, as well as a second year mathematics (Ordinary Differential Equations) class whose exposure to clickers was confined to a single activity. Participation was as follows:

1. A focus group of 6 (of 22) current clicker users from the 2011/12 Regression Modelling class
2. A survey of 7 (of 22) current clicker users – the same Regression Modelling class as the focus group (labelled 'STAT12' in the results)
3. A survey of 16 (of 24) previous clicker users – from the 2010/11 Regression Modelling class (labelled 'STAT11' in the results)
4. A survey of 32 (of 113) potential clicker users – from the 2011/12 ODEs class (labelled 'MATH' in results)

The clickers had been used on a weekly basis to answer course material related questions in the 2010/11 and 2011/12 Regression Modelling classes. For the ODEs class, the clickers had been used in one demonstration session only (based on answering a non-mathematical module evaluation questionnaire using a mobile phone application version of the Turning Point clicker software).

22.3 Evaluation

In total, 83% of students surveyed agreed that using clickers was or could be fun. The results are shown in Figure 22.1 below. Although numbers are small, it is possible that students who were most exposed to regular clicker use (i.e. the STAT11 and STAT12 students) were more likely to agree with this statement than those who had relatively little exposure to the clickers (i.e. the MATH group). This is consistent with other results in the literature that indicate that students like to use these devices.

We note that there is a "learning curve" for both lecturers and students associated with clicker use. Students were conscious of the time spent setting up and using clickers. Our experience, however, suggests that when used regularly this is generally not an issue. In the future we hope to further address this by

(a) issuing students with their own clickers,
(b) using them in most lectures and
(c) purchasing "newer" equipment that is easier to set up and use (from both perspectives).

We note that some feedback indicates that a few students find this process irritating. A few resent the "delay" induced by using clickers with a whole class.

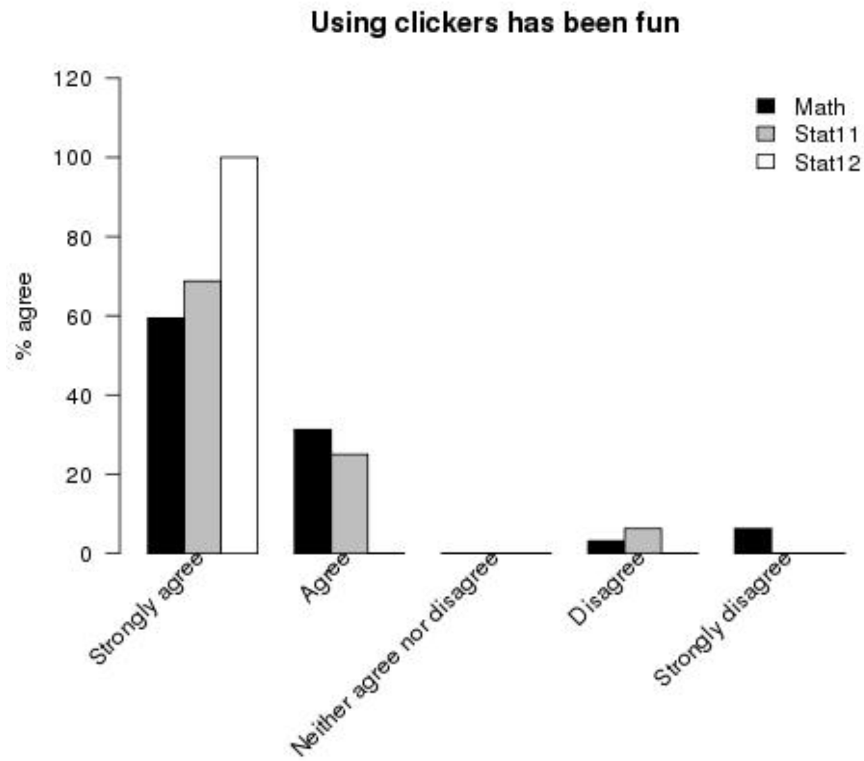


Fig. 22.1 Student response to question 1, "Using Clickers has been/could be fun".

However, the survey reinforces the comment that many appear to work more on problems when they are required to submit some answer via a clicker. The focus group indicated that some students can be reluctant to raise a hand in class; also some need more time to work through a question before they are in a position to raise a hand. We therefore asked in the survey whether using clickers rather than raising their hands gave more students the opportunity to work on problems and there appeared to be very strong support that clickers have value in this regard. Responses are shown in Figure 22.2. It is clear that the use of clickers was beneficial to all groups in encouraging participation in classes. Furthermore, it should be noted that there was only one person who disagreed with this statement from amongst the students who had regularly used clickers.

The response to negatively-phrased question 3, "I think that using clicker based exercises wouldn't really help me decide whether I understand something or not", is shown in Table 22.1 below. It indicates that those who are less familiar with the technology are less confident of its benefits.

We have been using clickers for several years, and have anecdotal evidence that at least some students find them useful. In this project, feedback from both the focus group and the survey provided plenty of evidence that a number of students find

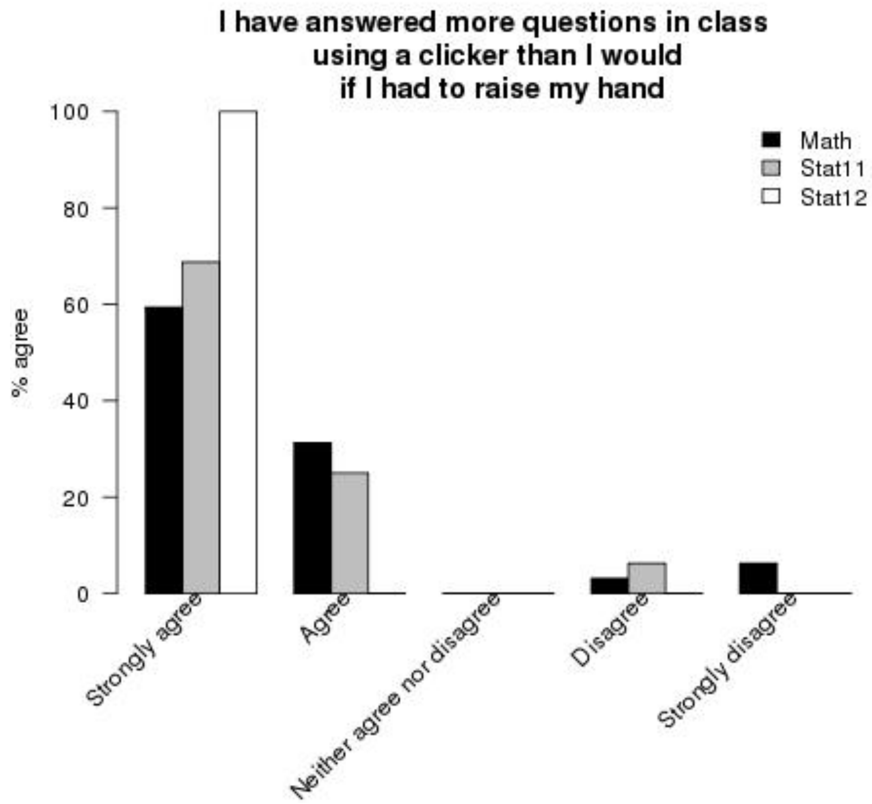


Fig. 22.2 Student response to question 2, "I did/would answer more questions in class using a clicker than I would if I were just asked to raise my hand".

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Math	7	8	11	5	1
Stat11	0	2	5	7	2
Stat12	0	0	1	4	2

Table 22.1 Student response to question 3, "I think that using clicker based exercises wouldn't really help me decide whether I understand something or not"

clickers useful in focussing their learning during lectures. This was evident from the response to question 4, shown in Figure 22.3 below.

Essentially, our focus group provided an indication that the use of clickers encouraged students to work on problems in class. The potential of clickers to foster active learning in maths/stats has been noted elsewhere (Kaplan, 2009) and the strong suggestion was that otherwise they may just take notes and only work on the material later (such as when cramming for exams). Clickers therefore appear to be a fairly gentle way of encouraging them to think about problems, check their under-

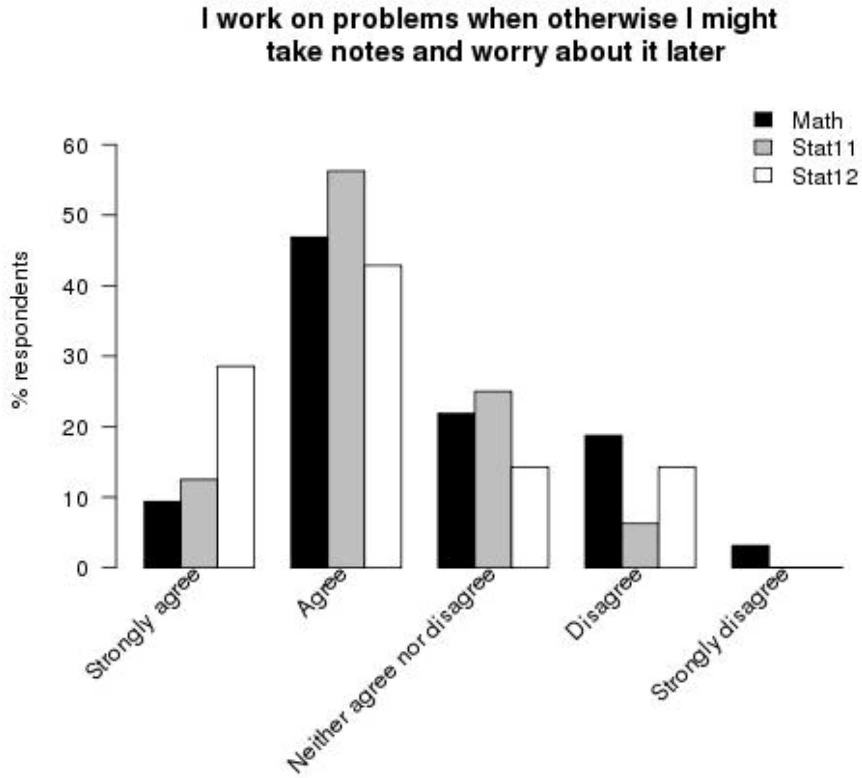


Fig. 22.3 Student response to question 4, "Using clickers would mean I would work on something when otherwise I might take notes and think about it later"

standing and generally engage with the material there and then. Consequently there was even some interest in getting records from clicker systems so that they could check their progress during the year and identify areas of strength and weakness.

A good example of the focus group feedback was as follows:

> Made you think and make a decision without risk of being shown up, encouraged you to try and work it out for yourself.

However, students were apparently not keen on their responses being made too public. Responses to question 5 and question 6 were intriguing and perhaps contradictory: clearly, it would be difficult for students themselves to keep track of completely anonymous results. However, perhaps the students are indicating mainly that their results should not be known to their peers in class. This is an aspect that should be investigated further.

There were also some interesting contrasts between student views on whether participation should be viewed as compulsory (see Table 22.4) and to what extent clickers could be used in formal assessment (see below), with the MATH group much less convinced that participation should be compulsory (and presumably less

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Math	16	8	5	0	3
Stat11	6	5	5	0	0
Stat12	4	1	2	0	0

Table 22.2 Student response to question 5, "I think clickers would work best if they were anonymous"

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Math	9	10	9	2	2
Stat11	8	4	2	2	0
Stat12	1	3	2	1	0

Table 22.3 Student response to question 6, "I would like to get my clicker results back so I can see what I'm good at and what I need to work on"

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Math	2	3	9	12	6
Stat11	0	4	10	2	0
Stat12	0	4	1	2	0

Table 22.4 Student response to question 7, "Participation in clicker exercises should be treated as an essential part of a module"

confident that merely participating in clicker exercises would be beneficial to them learning the subject).

22.4 Role in assessment

Finally, we explored in both the focus group and the survey the potential for using clickers in formal assessment. The MATH class was somewhat skeptical. In responding to question 8, "What would be your view if you were asked to complete in class tests using clickers? (You can select more than one answer)", 16 of the 32 students in this group would not want clicker-based in-class tests under any circumstances. None of the STAT students indicated this opinion. However, only one MATH, two STAT12 and one STAT11 student indicated that they would be unconditionally happy with any form of clicker-based assessment. This indicates two things of interest: that students may need to develop familiarity with and confidence in the technology, and that the provision of practice tests is essential. The responses to the second question on assessment, question 9, are outlined in Table 22.5 below.

What is not apparent from the summary table is the amount of complexity which lies behind the responses to this question. We assumed that the response to this question would be ambiguous: it had been clear from the focus group that pressures

	Number of agreements with this statement
Plenty of regular, low-stakes tests, formally assessed but worth very few marks	32
I would rather be marked for participation (and be able to learn from my mistakes without being marked down) than have the pressure of having to get it right each time	21
No formal clicker based tests under any circumstances	11
Only one or two formally assessed tests a term	6

Table 22.5 Responses to question 9, "If you were to have clicker based tests in a module, how would you expect it to work"

of having too much coursework was a concern of many students and some of the more positive views on assessment by clicker really seemed to tell us more about the need to offer a more realistic coursework burden. We even had one comment that told us that clickers would be good for learning but not for assessment, which also perhaps suggests we have some way to go in ensuring constructive alignment when offering coursework. We feel that it is possible to argue a case for the potential in developing summative assessment to encourage participation in large lectures by means of clicker-monitored problems and we believe the results above support this. Students were concerned with the limited assessment value of multiple choice questions. This is despite the fact we have been using clickers which offer more than just multiple choice. However, we accept that there are some problems that can be set on paper that cannot be set by clicker. So perhaps the role of clickers in summative assessment will be focussed on participation in lectures. Again, the focus group participants did express an appreciation of the way clickers can help them manage their learning.

22.5 Discussion, learning and impact

Careful consideration of our data suggests that there can be a role for clickers in maths/stats education. Although not considered here, clickers do also allow the lecturer instant feedback on student understanding and thus the direction of a lecture can be adapted as needed (Kaplan et al., 2008). As far as questions around summative assessment are concerned, student opinion seems divided but not particularly strong. In terms of depth of answer, it seems difficult to see what advantages clickers offer in summative assessment over other forms of assessment. There is, however, ample evidence that there are potential learning benefits when managing large classes. The unanswered question is whether a rubric is needed to encourage participation. One issue that did emerge is students' expectation of the role of lectures. For example we did get comments such as "I feel clickers would distract from the relaying of information from the lecturer's" as well as related comments indicating a preference for watching the lecturer "do" the examples. In order to more fully understand the potential role of clickers, clearly we need to better understand student expectations around learning and lectures.

22.6 Further development and sustainability

The School of Computing and Mathematics have purchased sufficient devices to equip the entire second year. Plymouth University has funded a Teaching Fellowship for the authors to further develop the work described here. This work will include pedagogic outputs as well as question banks developed from the material described elsewhere (GoodQuestions, 2012).

References

Bode, M. Drane, D. Kolikant, Y. & Schuller, M. (2009) A Clicker Approach to Teaching Calculus. *Notices of the American Mathematical Society*. 56, 253-256.

Chin, C. & Brown, D. (2002) Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*. 24, 521-549.

GoodQuestions. `http://www.math.cornell.edu/~GoodQuestions/`.

Kaplan, J.J. (2009) Promoting active learning in introduction to statistics using personal response systems (clickers). *Webinar, Consortium for the Advancement of Undergraduate Statistics Education (CAUSE).* `http://www.causeweb.org/webinar/teaching/2009-03/`

Kaplan, J.J. & Urban-Lurain, M. (2008) Personal Response Systems in Statistics: Using clickers to foster active learning and address student misconceptions. *Proceedings of the Inaugural Conference on Classroom Response Systems: Innovations and Best Practices*. `http://iclicker.com/dnn/UserCommunity/ConferencePapers/tabid/171/Default.aspx`

Kay, R. & LeSage, A (2009) Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*. 53, 819-827

Robinson, C & King, S (2009) Introducing Electronic Voting Systems into the teaching of Mathematics. *MSOR Connections.* 9(1), 29-33.

Chapter 23
Performance Assessment in Mathematics: Preliminary Empirical Research

Paola Iannone and Adrian Simpson

Abstract This paper outlines the experience of introducing a form of oral performance assessment into an undergraduate degree module. While oral assessment is commonplace in many countries, it has all but disappeared from undergraduate mathematics in the UK and we explore some of the issues regarding implementing this form of assessment, some of the potential advantages and how this particular form of oral assessment was used with a group of first year undergraduates. We discuss the outcomes in terms of students' performance and student and tutor views of the assessment process.

23.1 Introduction

The issue of how to assess undergraduate mathematics has been a significant one for decades. Most recently, Levesley (2011) noted a number of current challenges regarding assessment including

- the potential for conflict between mathematicians' ownership of assessment methods and the requirements of external quality assurance and management systems
- conservatism and risk aversion with universities
- the need to assess efficiently
- the need to avoid repetitive testing

Iannone and Simpson (2011) noted the assessment diet for students tends to be very restricted, with closed book examinations overwhelming all other forms of assessment. While most universities include elements of coursework, projects or

Paola Iannone, e-mail: p.iannone@uea.ac.uk
School of Education and Lifelong Learning
University of East Anglia

Adrian Simpson, e-mail: adrian.simpson@durham.ac.uk
School of Education
Durham University

dissertation assessment, the written examination contributes the vast majority of the marks towards the final degree classification.

In contrast, Stray (2001) noted that the oral examination was the norm until the start of the 20th century when it fell out of favour as a result of political issues and because written examinations were seen as more efficient. However, oral assessment in mathematics remains commonplace in many mainland European countries such as Italy, Hungary and the Czech Republic and there is evidence that they play a role in some US institutions, including as synoptic exams or oral assessments in group work (Gold, 1999).

We developed a project to explore how an oral assessment component might be introduced to a mainstream undergraduate mathematics module, by replacing one piece of coursework with a one-to-one tutorial in which students discussed the mathematics surrounding standard homework questions with a tutor.

23.2 Oral assessments - the research evidence

The concept of an oral component to assessment can cover a range of methods. Joughin (2010) contrasts three forms of oral assessment: *presentation* on a prepared topic (which may be individual or in groups); *interrogation* (from short form Q&A, through to the still ubiquitous doctoral viva) and *application* (where candidates apply their knowledge live in a simulated situation, e.g. when medical students diagnose an actor playing the part of a patient).

While presentations already play a role in many undergraduate mathematics degrees (Iannone and Simpson, 2011) our project was specifically focussed at an oral assessment of the second form. Rather than the pejorative term *interrogation* we call these *oral performance assessment*, since it requires students to work live on a problem with a tutor.

Joughin (1998) notes that oral performance assessments may have advantages in allowing probing of knowledge and more accurately resembling a real world problem solving situation in which one solves a problem in a dialogue with a colleague. Despite this, there is little literature examining oral assessment methods (Hounsell et al., 2007, note that only 2% of the works in their survey of assessment across all disciplines discussed oral performance assessment of this type).

Students' perceptions of oral assessment have been explored in a number of contexts. In social work, Henderson, Lloyd and Scott (2002) noted a marked difference between the generally negative views of them (particularly in relation to anxiety and usefulness) before they were undertaken, more positive ones afterwards and, once students had graduated and were practising professionals, a positive sense of their value and authenticity. Huxham, Campbell and Westwood (2012) similarly noted the issue of nervousness being tempered by an understanding of their usefulness and authenticity of oral examinations in a study involving biology students.

While there are some articles which discuss the implementation of oral assessment in mathematics (Gold, 1999; Nelson 2011), many of these are presentations and do not address the performance element of oral assessment. There is little empirical research in the use of orals in mathematics, their outcomes and student attitudes to them. This project was designed to examine the implementation of one particular variant of oral performance assessment in an undergraduate mathematics module and describe the student and staff experiences of that implementation.

23.3 Pre - implementation issues

While many mathematicians express concerns about the ways in which undergraduate mathematics is assessed (LMS, 2010) and oral performance assessments have been proposed as potential solutions (Levesley, 2011), one area we were interested in was the practicalities of implementation and we did find some obstacles to implementing oral performance assessment. While all the staff involved were willing and, in many cases, generous in giving time to the project, there was a nervousness about permitting this form of assessment. These seemed to focus around four main areas: institutional constraints, fairness, anxiety and preparedness.

Some people involved were unsure where the authority lay to vary the assessment procedure or whether it would be acceptable to the institution. This may have been related to the short timescale in which the project was implemented, requiring making a minor variation to the assessment process during the academic year. The inertia in systems governing teaching and assessment practices in many universities can appear to lead to delays of years to get even minor changes approved and can stifle change (Bryan and Clegg, 2006). In our case, because the change was minor and the timescale tight, approval was given from the appropriate committee chairs in the University outside the usual change routes. There appeared to be no real barriers to implementation and some genuine interest was expressed in the outcomes of the project.

As we prepared the assessment, a number of people raised the issue of fairness. By their nature, oral performance assessment (as with other performance assessments such as music and drama performance, oral language tests, driving tests, etc.) cannot be truly anonymous and this can give rise to concerns about bias. However, one could argue that the potential for bias is impossible to eliminate from any assessment system: it is possible that one can even be biased by handwriting style with anonymised written examination scripts (Briggs, 1980). Thus the issue of fairness is really one of monitoring and moderation: if there are concerns about the fairness of marking with written work, it can always be re-examined and assessment procedures often undergo moderation to reduce any potential for bias. While having a second assessor in each tutorial was considered, given the existence of cheap, high quality video recorders, we instead opted to record the tutorials to enable marks to be challenged and moderated (with the understanding that they would be deleted at the end of the assessment process unless students explicitly agreed to their use for research). Another issue of fairness raised concerned students with English as an additional language. However, we felt that difficulties with English would be equally likely in written submissions and the opportunity to help students express themselves orally might be easier in a suitably sensitive tutorial setting.

A number of people voiced concerns about the level of anxiety a one-to-one tutorial might cause. However, the regular tutors for the module noted that their previous encounters with students on an individual basis had shown they were more comfortable with direct conversation than speaking in a tutorial setting with other students present. This fits with the evidence from Marshall and Jones (2003) that while anxiety was higher for oral clinical examinations than written examinations, it was higher still for seminar presentations.

The anxiety level of oral performance assessment may be higher than that for written assessment partly because students have considerably less experience of them. While we had no opportunity to give students the chance to practise this form of assessment, as much as possible was done to deal with this. Assessors were asked to make the setting and initial part of the conversation as comfortable as possible, students were given a short talk about the planned assessment which emphasised the conversational nature of the tutorial and this was re-emphasised in the information about the organisation of the tutorials.

23.4 Implementation

The assessment took place in a first year module on graph theory. While not a compulsory module, it is taken by the majority of first year students, with 108 students registered. The marks for the module are made up from 10% for solutions to homework problems and 90% for a written examination taken at the end of the year. The oral performance assessment replaced one set of homework problems and took place during normal tutorial time (which in other weeks would have been used to discuss and return the homework problems).

Each student was asked to attend a 10 minute session described thus:

> The idea of the one-to-one tutorial is to help you express what you understand, not to catch you out. It is perfectly acceptable to ask the tutor for help or for the tutor to give you guidance or to help correct any errors you've made which might make it difficult for you to get to the answer. You'll be able to use the blackboard or paper to write things down (though you shouldn't bring complete answers or notes with you - we want to talk to you about the problem and its solution, not just hear you read an answer out!).

As normal, they were set the homework problems during the previous week, so they had at least six days to work on them. The problems were the following:

A: Prove that if a graph has at least 11 vertices, then either it or its complement must be non-planar.

B: Show that every connected planar graph with less than 12 vertices has a vertex of degree 4 or less. [Hint: argue by contradiction to get a lower bound for the number of edges which contradicts the upper bound which follows from Euler's formula]

C: For each graph find a minimum spanning tree and prove it is unique:

(a) Q_3 with the usual binary vertex label and weigh $w(ij) = i + j$.
(b) K_5 with vertices $\{1,\ldots,5\}$ and weigh $w(ij) = i + j^2$ where $i < j$.

D: Draw all forests on 5 vertices and justify your answer.

The students were told that in the tutorial, of the four problems set, they could choose one to discuss first and then the tutor would choose another. Because two of the questions required a proof (A and B) and two required the use of an algorithm and some reasoning about the outcome (C and D), it was agreed that when the student chose a question from one pair, they would be assigned a question from the second pair through some random process (tossing a coin or drawing lots). Students were also informed that they would be videoed to allow for marks to be moderated.

Tutors discussed the idea of *contingent questions* – areas they could explore around the solution, depending upon the quality and form of the response the student gave. For example, in question D, depending on the kind of method employed, the student could be asked to explain how they would check to ensure that no two forests were isomorphic, how they would prove that their method gave an exhaustive list or how they might use their solution to estimate the number of forests on 6 vertices.

The tutorials normally took place in groups of 12-16. Five people acted as assessors (the course lecturer, two postgraduate tutors who normally ran the tutorials and the two authors). For each tutorial slot, four assessors were needed. However, since no extra marking was required, the resource overhead was not as large as it would first appear. The 16 hours of staff time normally used for marking work and delivering tutorials were replaced with just over 18 hours of staff time doing the one-to-one tutorials.

At the end of each 10-minute slot, the tutor awarded a mark based on an assessment matrix (Figure 23.1) which had also been shared with the students.

The week after the assessments were completed, students were asked to fill in a short questionnaire adapted from the Assessment Experience Questionnaire (AEQ from Gibbs and Simpson, 2003) comparing the one-to-one tutorial with written coursework. They were also all contacted to see if they were prepared to attend a short interview to discuss the assessment and their experience of it. The course lecturer and the two postgraduate students involved in conducting the assessments were also interviewed.

23.5 Outcomes

Each question was marked out of 5 and the performance was generally good. The mean marks for each question A, B, C and D were 4.00 (σ = 0.93), 4.16 (σ = 1.09), 4.06 (σ = 1.01) and 3.89 (σ = 1.18) respectively (with no statistically significant differences between any pair of questions) and the students averaged 7.96 across the

Grade	Solution	Key ideas and application	Clarity and explanation
5	Complete solution outline given with no extra help needed	Clearly identified key ideas behind the problem and shown how they apply elsewhere	Explains clearly and concisely, even in unfamiliar areas
4	Complete solution given with some extra help	Identified key ideas or shown how solution approach might apply elsewhere	Explains clearly and concisely in prepared areas and generally clear elsewhere
3	Complete solution given with substantial extra help	Has identified some key ideas, but may not fully distinguish key ideas from calculations or details OR shown some sense of wider application of solution	Explanations need a little probing to clarify
2	Complete solution not obtained, but some key steps made without help	Does not have key ideas or any sense of wider application	Explanations need to be drawn out at length
1	Complete solution not obtained, but some key steps made with help	Does not have key ideas or any sense of wider application	Has difficulty giving any explanations

Fig. 23.1 Assessment Matrix

two questions they were asked. Over the previous 6 weeks of coursework (covering the graph theory component of the course) the average mark was 4.29 ($\sigma = 0.95$).

While the marks obtained for a given tutorial are highly dependent on the choice of questions, it is worth noting that the one-to-one tutorial marks and their spread seem broadly similar to those for previous written homework. The attendance was also similar to previous weeks: of the 108 students registered on the course, only 9 failed to attend their assigned one-to-one tutorial. This compares with an average of 8% each week failing to attend a tutorial and 9% failing to submit the homework.

The issue of anxiety is one which pervades the literature on oral assessment and had been a concern in planning, but only one student expressed any concerns before the assessment. One further telling indicator that the level of anxiety was much less than that supposed before implementation came from the administration of the videos. To comply with ethical research standards, students were asked to give their consent for the use of the videos for research purposes: if students were unduly nervous about being videoed one would have expected few to opt in. In fact 97 out of the 99 students agreed to their use.

Figure 23.2 shows the mean (with standard error bars) of the students' scores for each statement in the AEQ. They were asked to rate each statement as being more accurate of weekly sheets or more accurate of the one-to-one tutorials (on a five point Likert scale from +2 to -2). The AEQ proved highly reliable (α = 0.81) and while space precludes a full analysis of the results, there were some interesting and surprising findings worth reporting.

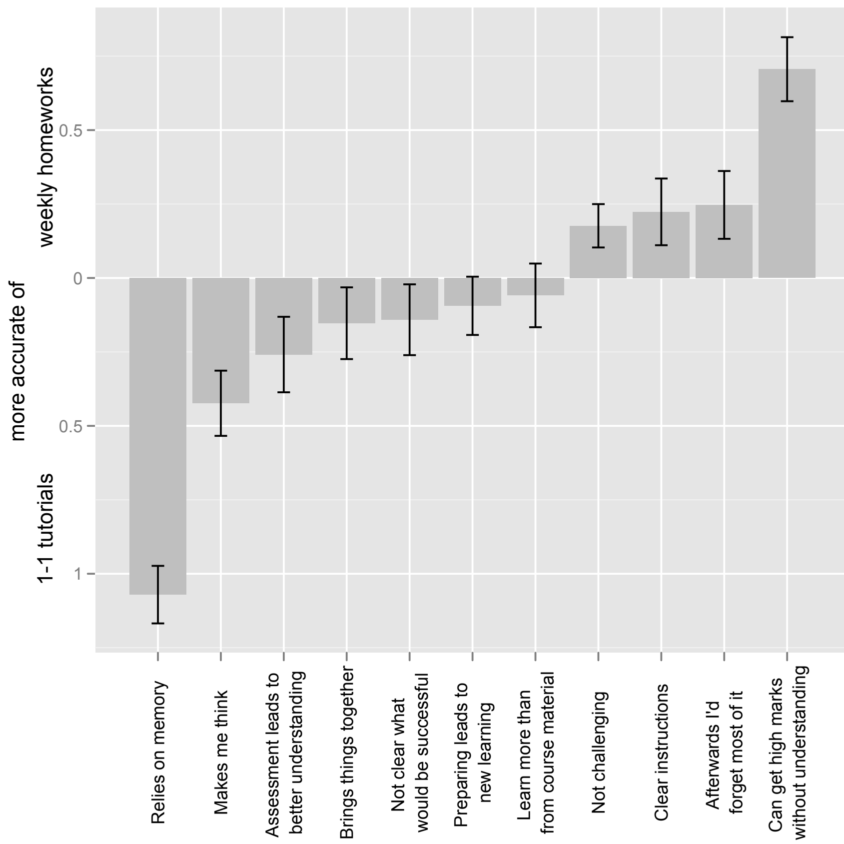


Fig. 23.2 AEQ Results

The students saw the one-to-one tutorial as making them think significantly more about the material ($t(84) = 3.84$, $p < 0.001$) and helping them understand things significantly better ($t(84) = 2.02$, $p < 0.05$) than weekly example sheets. They felt that with the weekly example sheets it was easier to get away with not understanding and still get high marks ($t(84) = 6.52$, $p < 0.001$), that they were not as challenging ($t(84) = 2.41$, $p < 0.05$) and they were significantly more likely to forget the material learned ($t(84) = 2.16$, $p < 0.05$) than one-to-one tutorials. However, they also thought that the one-to-one tutorial relied more on memory than example sheets ($t(84) = 11.00$, $p < 0.001$).

In addition to the comparative questions, students were given some free text to discuss their experience of the one-to-one tutorials and many took the opportunity to give substantial responses. While, again, space prevents a detailed analysis of these, the general tone came across clearly through repeated comments. Many mentioned that they had prepared more than they would for an ordinary weekly homework and tutorial and felt that they needed to understand the material more deeply to be successful. However, many also mentioned that they found the situation more stressful and the pressure of the short timescale only added to this. Some commented on the need to rely on their memory more than for written homework, but many noted that there was value in individual working and feedback tailored to the gaps in understanding.

Many mentioned the issue of fairness and consistency, particularly in relation to the contingent questioning. Some felt that the tutors being able to tailor questions to their solutions was a benefit of the system, but others were concerned that not everyone got identical questions. Quite a lot of students noted both benefits and disadvantages and the general sense was that the oral performance assessment could have a role to play in a more mixed diet of methods, alongside (rather than replacing) written example sheets and exams.

In the interview with the course tutors, some of the advantages and drawbacks of this type of assessment were discussed. The tutors strongly agreed that performance assessment of this kind allows the assessor to find out 'what they [*the students*] do know rather than what they don't know' and allows the assessor to get to know the students individually and quickly understand whether they are struggling, or indeed whether they are coping well and on top of the material. A setting like the one-to-one tutorial does not allow a student to 'hide' amongst his/her peers and lets the assessor offer targeted help. Concerns were raised about accountability (although the assessors acknowledged that the marks could be moderated by using the videos) and resources. On the whole the tutors considered this as a positive experience, but felt that much more work went into this type of assessment than in the usual marking of weekly exercise sheets. However, this may have been because they were comparing the effort related to preparing and piloting a new form of assessment with a regular, well systematised form: the actual time spent in assessing was broadly similar to the time given to a week's tutorials and marking.

In conclusion, there were a number of concerns raised during implementation: many students reported levels of anxiety, though generally the students reported a positive experience. The workload and the results in terms of attendance and marks were broadly similar to the ordinary coursework/tutorial system, but the students and staff reported that the assessment led to improved understanding, was less likely to enable students to gain high marks without understanding and was more likely to engage the students in thinking about the material.

References

Briggs, D. (1980) A Study of the influence of handwriting upon grades using examination scripts, *Educational Review*. 32(2), 186-193.

Bryan, C. & Clegg, K. (2006) *Innovative Assessment in Higher Education*, Routledge.

Gibbs, G. & Simpson, C. (2003) Measuring the response of students to assessment: the Assessment Experience Questionnaire. 11th *International Improving Student Learning Symposium*, Hinckley, England.

Gold, B. (1999) *Assessment Practices in Undergraduate Mathematics*, Washington DC: Mathematical Association of America, 143-145.

Henderson, J., Lloyd, P. & Scott, H. (2002) 'In the real world we're all put on the spot at some time or other, so you need to be prepared for it': An exploratory study of an oral method of assessing knowledge of mental health law. *Social Work Education*. 21(1), 91-103.

Hounsell, D., N. Falchikov, J. Hounsell, M. Klampfleitner, M. Huxham, K. Thompson, & S. Blair (2007) *Innovative assessment across the disciplines: An analytical review of the literature*. York: Higher Education Academy.

Huxham, M., Campbell, F. & Westwood, J. (2012) Oral versus written assessments: a test of student performance and attitudes, *Assessment & Evaluation in Higher Education*. 37(1), 125-136.

Iannone, P. & Simpson, A. (2011) The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*. 30(4), 186-196.

Joughin, G. (1998) Dimensions of oral assessment. *Assessment & Evaluation in Higher Education*. 23(4), 367-78.

Joughin, G. (2010) *A short guide to oral assessment*, Leeds Met Press.

Levesley, J. (2011). Taking control of the assessment agenda. In Rowlett, P. (ed) *Report of the HE Mathematics Curriculum Summit*, York: Higher Education Academy, 21-23.

London Mathematical Society (LMS). (2010). *Mathematics Degrees, their Teaching and Assessment*. Accessed via `http://futurelms.wordpress.com/2010/04/08/teaching-position-statement/`

Marshall, G. & Jones, N. (2003) A pilot study into the anxiety induced by various assessment methods, *Radiography*. 9, 185-191.

Nelson, M. (2011) Oral assessments: improving retention, grades and understanding, *PRIMUS*. 21(1), 47-61.

Stray, C. (2001) The shift from Oral to Written Examination: Cambridge and Oxford 1700-1900. *Assessment in Education*. 8(1), 33-50.

MU MAP was supported by the MSOR Network through the Mathematical Sciences HE Curriculum Innovation Project, part of the National HE STEM Programme.

This book discusses the outcomes of the MU-MAP Project (Mapping University Mathematics Assessment Practices) aimed at detailing the current state of assessment practices in undergraduate mathematics including:

- *A survey of existing practices at universities across England and Wales*
- *A summary of the research literature*
- *Examples of different forms of mathematics assessment in current use*
- *Reports on the implementation of changed assessment projects such as oral assessment, the use of applied comparative judgement techniques and assessing employability skills*

More information about the outcomes of the MU-MAP project is available at www.uea.ac.uk/edu/mumap

ISBN 978-1-870284-01-1
90000 >
9 781870 284011