CULMS Newsletter

Number 8

December 2013

Community for Undergraduate Learning in the Mathematical Sciences

The CULMS Newsletter

CULMS is the Community for Undergraduate Learning in the Mathematical Sciences.

This newsletter is for mathematical science providers at universities with a focus on teaching and learning.

Each issue will share local and international knowledge and research as well as provide information about resources and conferences.

Supported by:
New Zealand Institute of Mathematics and its Applications (NZIMA):

Editors:	Bill Barton, The University of Auckland Greg Oates, The University of Auckland Mike Thomas, The University of Auckland
Published by:	The Mathematics Education Unit, The University of Auckland, Private Bag 92019 Auckland, New Zealand
Correspondence:	Greg Oates g.oates@auckland.ac.nz
Website:	www.math.auckland.ac.nz/CULMS

Previous editions of the newsletter can downloaded from our website or for an electronic copy, please email Greg Oates at g.oates@auckland.ac.nz

ISSN 2230-3804 (Print)
ISSN 2230-3812 (Online)

CULMS Newsletter

Number 8 December 2013

Foreword 2
Greg Oates

Opinion Piece: Unwritten Contracts 4
David Holgate

Progress on Improving Student Success in First-Year Mathematics at the University of Canterbury 7
Liz Ackerley & Rua Murray

Can Students Tackle a Big Problem When Time is Short? 11
Peter Bier

Capturing Undergraduate Learning in the use of STACK in Mathematics at Birmingham 2012–2013 16
C. J. Sangwin & D. F. M. Hermans

Foreword

This eighth edition of the CULMS newsletter is the second published in electronic format in 2013, available for download from our website at:

www.math.auckland.ac.nz/CULMS

This edition is timely, given that it follows soon after the 9th biennial Delta Conference on the teaching and learning of undergraduate mathematics and statistics, held in Kiama New South Wales from the 24th to the 29th November 2013. Like the papers presented at the Delta conference, the papers in this 8th *CULMS* newsletter highlight significant contemporary issues and challenges facing the undergraduate sciences, from delivery and assessment, through to the interactions between academics and their students. All the authors in this edition have also been regular contributors to the Delta symposia, and further studies by two, David Holgate and Peter Bier, may be found in one of the two publications produced by the Delta 2013 organisation. The first of these is a special edition of the journal *iJMEST* (Volume 44(7), October 2013), guest-edited by Kelly Matthews, with 11 papers covering a broad range of issues. The second, published in electronic format for distribution to conference delegates, contains a further 21 papers and 50 abstracts which gives some appreciation of the breadth of topics covered in the conference presentations (*Shining through the fog: Proceedings of Lighthouse Delta 2013, The 9th Delta Conference of teaching and learning of undergraduate mathematics and statistics,* edited by Deborah King, Birgit Loch and Leanne Rylands). We hope this latter publication will soon be available for all *CULMS* subscribers to download from the Delta website which is currently under construction.

The first article in this 8th *CULMS* Edition is a candid opinion piece by David Holgate, the Head of the Mathematics Department at the University of the Western Cape in South Africa. David ponders the nature of the relationship between academics and their students in a way which we are sure will resonate with many readers, and suggests possible ways in which explicit recognition of the nature of these relationships may be productively used. Liz Ackerley and Rua Murray describe the way in which they have attempted to better match students to the appropriate course in their university, and provide better and more meaningful support for students. Again, the situation they describe should resonate with readers, and their solution seems eminently feasible for those wishing to emulate some of their efforts. Peter Bier describes an ambitious attempt to implement complex-problem-based in-course projects within the confines of a short summer-school engineering course, and his paper details both the creative nature and the success of his efforts which have been widely recognised within the Engineering Science domain. Chris Sangwin and Dirk Hermans describe the long-term project they have implemented at the University of Birmingham, another ambitious project which has again received wide international recognition. Their *STACK* project allows for CAS-based online assessment which moves beyond the common static Multiple-Choice or True-False style questions, to include numeric and algebraic input of student solutions. It is almost certain that STACK allows for the most complex level of student-input and evaluation of the subsequent responses of any such other system currently available, as can be seen from the examples they include in this article.

We hope you find the articles in this edition stimulating and interesting, and we encourage you to follow up further with the Delta publications cited previously.

Note: We are continually looking for interesting and relevant submissions that consider new developments, research and practice in the teaching and learning of undergraduate mathematical sciences, including those that address the transition from secondary to tertiary levels. Please email submissions to:

Greg Oates, Department of Mathematics, The University of Auckland
Email: g.oates@auckland.ac.nz

Unwritten Contracts

David Holgate,
University of the Western Cape, South Africa

I recall with some embarrassment a reading course I once gave to a graduate student. We had a programme planned for the semester built around a progression of exercises. On agreed dates he was to submit solutions to the exercises and in between we would meet to discuss the reading and I would give him feedback on his solutions. Two thirds of the way through the semester our course ground to a halt.

In the words of my student, we had reached a "ceasefire". He owed me a set of exercises and I had yet to mark his previous submission. We were both preoccupied with other obligations and neither of us felt morally able to insist that the other uphold his end of the contract. In truth it was a situation which suited both of us quite well, each being absolved to some extent by the inactivity of the other.

This scenario caricatures an interaction between student and academic with which, I am sure, we can all identify. There are times when we are only too happy for our students to leave us alone, and they in turn would prefer to have academic demands under their control not ours.

The Disengagement Compact

George Kuh (1991) in his work on the student experience has highlighted and challenged this interaction between students and academics at an institutional level. He coined the term "disengagement compact" to describe the unspoken agreement which often exists between faculty and students, keeping a safe distance from each other so that all parties can meet external expectations while minimising mutual inconvenience.

While the "ceasefire" with my graduate student is a good local illustration of the outworking of such a compact, Kuh's notion is broader and less frivolous. He is challenging an institutional culture and asks us to examine how university policies, and in particular performance measurement, can have unintended impact on our core activities and thus the quality of education we offer.

When I first encountered the notion of a "disengagement compact" it struck an immediate chord. Certainly that was in part because it challenged me to examine my own interaction with my students. More fundamentally, however, naming and making explicit a universal yet unspoken code of conduct not only moves us to self-examination but opens the opportunity for broader collegial catharsis, ownership of our joint professional development and a more fruitful teaching and learning environment.

Take Explicit Responsibility For Our Own Development

A quick internet search will reveal widespread comment and reflection on Kuh's "disengagement compact". Many respond with fatalism to the inevitability of the reality. Much of the writing is of course directed at the institutional level. My intention is not to reflect on Kuh's writing *per se*. Despite institutional pressures and constraints we still enjoy enormous freedom as university academics which we can exploit for our personal and corporate professional development. With this in mind I challenge us all to make explicit the unwritten and unspoken "contracts", "treaties" or "understandings" which exist between student and lecturer.

Holgate, D. (2013). Unwritten Contracts. In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS)*, ***8***, 4-6.

Consider the extent to which you are complicit in such pacts. Think about how you can address the underlying issues with your students and colleagues. Teaching is at heart a relationship and no relationship can thrive on unspoken expectations and misunderstandings.

What Contracts Are You Party To?

To start the ball rolling, I offer some “unwritten contracts” which I have observed in my teaching experience (A number of them can be viewed as corollaries of the disengagement compact, I guess). I have found it helpful to consider how they impact on my teaching and on my students’ learning. You can no doubt identity others.

I am a paying client; it's your job to get me through:

This “contract” is usually invoked at times of failure, by students and on occasion by parents too. It reveals an attitude that should be addressed long before failure occurs and also challenges us to consider what our role is in a more commercialised, contractual higher education environment. (This contract is closely related to the “*You can't fail me, I need this course to be able to carry on with my degree*”-contract, although the latter is more an appeal for pity.)

I need you to pass, my promotion depends on it:

(Alternately: You need me to pass, your promotion depends on it.) Like most unwritten understandings, this calls to question responsibility and motivation in the teaching-learning relationship. Of course we all experience pressure to pass (enough) students. To what extent does our students’ awareness of this impact their engagement? Can it lead to an instructive discussion of mutual responsibility for learning?

If we keep quiet long enough you'll answer the question:

Perhaps more a strategy than a non-verbal understanding, this can surely more fruitfully lead to an explicit conversation than a pattern of classroom interaction?

The expert is entitled to be lazy:

I vividly recall a lecturer in my undergraduate years who relied on this agreement. He had authored the textbook prescribed for the course which clearly excused him from preparing for lectures! We all fall back on this contract to some extent, yet no doubt despair when our talented students appropriate it for their own use.

I am sure you can add your own examples of agreements, understandings and misconceptions which are never verbalised and yet have significant impact on the quality of the teaching and learning we strive to foster. They can direct our students’ educational experience and impede our own professional development.

To return to my opening anecdote; the moment my student explicitly named our impasse as a “ceasefire” opened the opportunity for us to discuss our mutual responsibility, take action and complete the course. Identifying and putting into words the unwritten contracts which exist between ourselves and our students can be challenging and liberating and enhance our development as academic professionals.

Reference

G. D. Kuh, J. H. Shuh, E. J. Witt & associates (1991). *Involving colleges: successful approaches to fostering student learning and personal development outside the classroom.* San Francisco: Josset-Bass.

Author
David Holgate, University of the Western Cape, South Africa.
Email: dholgate@uwc.ac.za

Progress on Improving Student Success in First-Year Mathematics at the University of Canterbury

Liz Ackerley & Rua Murray
Department of Mathematics and Statistics, University of Canterbury, New Zealand

Here at UC we are wrestling with a persistent, and seemingly ubiquitous, issue for university mathematics programmes: W*hat should we do to improve success in first year courses?* Our goals are in fact quite specific; we are focused on first semester courses which take students directly from high school, and by *success* we mean that students should achieve a passing grade which allows progression into further courses. We are inspired to write this note in part by reading the contributions of Lenz (2013) and Begg (2013) in the previous CULMS newsletter, and partly to report the success that one of us (LA) has had in improving pass rates in the first course MATH102. Our experiences may be of interest to CULMS readers because the changes made to MATH102 are extremely saleable to colleagues: the curriculum is almost unchanged, the difficulty of final assessments is unchanged, and the delivery style (large class lectures plus small group tutorials, with some online activities) has persisted unscathed. What we have done is to *provide a whole package* of early preparedness assessment, catching failure early and ongoing, persistent support.

Each year, between 1000 and 2000 individual students take first year mathematics courses at the University of Canterbury; all intending engineering, economics, mathematics, physics and statistics students must enrol in our courses. Engineers take different courses from science and commerce students. Students who arrive with fewer than 18 units of level 3 NCEA Calculus are *advised but not required* to take the preparatory course MATH101 (no calculus knowledge is assumed). Any student may enrol in the advancing course MATH102 (linear algebra and calculus) and engineering students are required to take the similar course EMTH118. Bundled together, typical pass rates for the Semester 1 cohort of students have been around 65% (see Table 1), with variability between course occurrences.

NCEA Calculus Credits Achieved is a Relatively Weak Predictor of Student Success

Although NCEA achievement in various subject domains is positively correlated with first year University achievement (Brogt, Sampson, Comer, Turnbull & McIntosh, 2011), prediction of first year success on the basis of NCEA credits alone is not entirely reliable. For example, exclusion of students with fewer than 18 NCEA level 3 calculus credits would improve pass rates to as high as 84% (James, Montelle. & Williams, 2008), yet many students pass without this level of attainment. Consequently, our department does not presently require a certain number of NCEA credits as a prerequisite for course entry.

Ackerley, L. & Murray, R. (2013). Progress on improving student success in first-year mathematics at the University of Canterbury. In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS) Newsletter*, ***8***, 7-10.

Table 1. Pass rates of Semester 1 Mathematics courses at the University of Canterbury. Data are aggregations of MATH101, MATH102, EMTH118 (MATH108 before 2010). Notes: a) 2011 Semester 1 exam period was disrupted by the June 13 Earthquake, and 213/992 students received an adjusted grade; b) grade scale changes in 2012 reduced the use of 'non-advancing pass' grades.

YEAR	Clear Pass	Non-advancing pass	Failing grades
2008	64%	8%	28%
2009	64%	12%	24%
2010	63%	12%	26%
2011[a]	71%	3%	26%
2012[b]	65%	1%	35%

MATH102: A Course 'Out of Line'

Our semester one advancing course MATH102 has been identified as having negative student grade variability over the last three years: on average, students in Math102 have scored 1.1 grade points *below* their GPA for that semester[1]. This pattern repeated in second semester occurrences, with overall pass rates around 54%. About 25% of the students attained E grades, and most of these students were not sitting the final exam or had an established pattern of non-participation. LA was asked to coordinate the first semester occurrence of MATH102 in 2013 with the brief: "MATH102 does not seem to be working well for either the students in the course or the Department. Try to do something about this."

Rescuing MATH102

Early intervention was needed to address non-participation and turn the course around. The strategy we settled on was to have more contact between the course coordinator and the students in those vital first two weeks. The MATH102 teaching staff needed to spend more time focusing on the needs of the students.

Strategy I: MATH101 or MATH102?

The main aim in the first two weeks was to identify the weaker students in MATH102 and move them into our introductory mathematics course, MATH101. As noted above, this was not done via filtration on the basis of NCEA results. The first tutorial was scheduled in week one (not later) and was held in the labs to familiarise students with the software package used for our online quizzes (MapleTA). In this lab students worked their way through review MapleTA units. This had the added benefit of preparing them for an early diagnostic test (see below). LA (course coordinator) went to *every* lab and chatted to each of the (approximately) 160 students. The aim was to find out why they were doing MATH102, what their background in maths was and how they felt about the course so far, and then to advise them accordingly. LA developed a diagnostic online MapleTA test worth 3%, with randomised questions. This was based on background material and was open for 24 hours early in week two. Students were able to attempt the test as many times as they wished, and their best attempt was counted. (Note that once such a test is written, administration associated with resits, marking and recording of data

[1] The UC grade point scale ranges from -1 (E) to 9 (A+).

typically takes about 10 minutes total.) Completion of this test was a course requirement and students were informed that they needed to be scoring at least 80% to be confident of coping with the demands of the course. If they failed to achieve this level of competency then follow-up action would be taken. This was done via email (not effective), talking to students in lectures (a little more effective), phoning them (very effective, but time consuming) and working with the commerce and science advisors (also very effective and time consuming).

Strategy II: focused intervention

The mid-course test was held in week six and the borderline students (those scoring between 35% and 60%) were emailed warning them they were at risk of failing. They were also told that daily extra help sessions were available. Phoning students (or text messaging) would have been preferable but there were time limitations. Next year, more administrative support is planned for overseeing and coordinating these activities.

Outcomes

21 students moved from MATH102 into MATH101 in the first two weeks, and 10 students discontinued after the online test in week two. Student participation in lectures, quizzes and tutorials was high in term one and there was an 88% pass rate in the mid-course test. Participation and results dropped off in term two, but even so, only 9 of the 156 remaining students did not sit the final exam and an overall pass rate of 79% was achieved.

We were very pleased with the significant improvement in student outcomes achieved by this series of 'small steps'.

Acknowledgements:

Irene David, Alex James and Phillipa Williams also contributed to a Departmental working group assembling recommendations on which the improved outcomes were based.

References

Begg, A. (2013). Mathematics-Bridging, 101, 201, or what? In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS) Newsletter*, ***7***, 13-20

Brogt, E., Sampson, K.A., Comer, K., Turnbull, M.H. & McIntosh, A.R. (2011). Using institutional research data on tertiary performance to inform departmental advice to secondary students. *Journal of Institutional Research 16*(2), 26-41.

James, A., Montelle, C. & Williams, P. (2008). From lessons to lectures: NCEA mathematics results and first-year mathematics performance. *International Journal of Mathematical Education in Science and Technology (IJMEST)***,** 39(8), 1037-1050.

Lenz, K.E. (2013). Entry-level mathematics placement study at the University of Minnesota Duluth. In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS) Newsletter, 7*, 7-12.

Authors
Liz Ackerley, Department of Mathematics and Statistics, Te Whare Wānanga o Waitaha, University of Canterbury, New Zealand
Email: liz.ackerley@canterbury.ac.nz

Rua Murray, Department of Mathematics and Statistics, Te Whare Wānanga o Waitaha, University of Canterbury, New Zealand.
Email: rua.murray@canterbury.ac.nz

Can Students Tackle a Big Problem When Time is Short?

Peter Bier
Department of Engineering Science, University of Auckland

How far do you have to dive underwater to be safe from gunfire? That was the question I posed to my first year mathematical modelling class at the start of one Wednesday in summer school. The students then spent the next seven hours answering that question, in groups of three or four. Tackling this problem required researching relevant information, drawing on a wide range of mathematical skills to model the situation and producing a high quality written report. Is this too much to ask students to achieve in a short space of time? All but one group successfully rose to the challenge (and they failed due to problems with group dynamics).

Why a Day-Long Group Project?

Incorporating a day-long group mathematical modelling project into *Mathematical Modelling 1*[1] was born out of a desire to expose students to challenging real-life applications of mathematics, as opposed to the smaller "bite-size" examples we often deal with in lectures and tutorials. Tackling the complexities of an open ended modelling problem can be difficult in the lecture environment. Time is short, which requires examples to be small enough that they can be worked through in an hour or less. Since repetition of a few examples is desirable it is rare that students meet a problem in class that cannot be worked through in fifteen to twenty minutes.

Individual project work is one way expose students to problem-solving on a larger scale but I have found it comes with several major problems, particularly when coupled with the shorter time span of a summer school environment. In our first year modelling paper we eventually discarded individual project work, in part because it was known a large amount of collaboration was taking place. Marking a high volume of individual projects in a timely and consistent manner was also difficult, requiring a team of markers and a highly prescribed mark schedule. This in turn forced students to follow a particular approach to the problem, denying the very creativity I wanted to foster and encourage. Furthermore summer school lacks the two week mid-semester break, which has traditionally been the time period used for project work.

Moving to a day-long group project dealt with several major problems, while also building in some other very desirable outcomes. Collaboration was no longer an issue, as team-work was now an integral part of the assessment. By moving to group work the volume of marking was quartered, allowing one experienced person to mark all scripts (ensuring consistency and allowing for creative approaches to solving the problem). Rather than the time constraints of summer school teaching being a problem, they became part of the solution, with students learning to work as a team under pressure. In addition students were now exposed to research and mathematical report writing. These benefits came at the cost of only one day of teaching being lost.

[1] Mathematical Modelling 1 (ENGSCI 111) is the standard first year mathematics paper taken by engineering students at the University of Auckland. It covers mathematical modelling and techniques (including calculus, linear algebra and probability).

Bier, P. (2013). Can students tackle a big problem when time is short?. In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS) Newsletter*, ***8***, 11-15.

The idea of moving to a day-long group modelling project was not born out of a vacuum. For the last five years I have been involved with administering a day-long group mathematical modelling competition for senior secondary students (NZ's Next Top Engineering Scientist). This nationwide competition has been extremely successful, growing exponentially in popularity. Teachers frequently comment at how much students enjoy the event and how much they learn from it. The high quality of the competition entries opened my eyes to just how much can be achieved by a group of students in a short space of time.

Practicalities

The choice of group size was led by the success of NZ's Next Top Engineering Scientist where students work in teams of three or four. This is in line with the recommendations by the *Teaching and Engineering Education in Europe Thematic Network,* which suggests that the ideal group size for project based learning is between three to six students (Borri & Maffioli Firenze, 2007). It is also consistent with published results which indicate that groups of four or five members work best for project based group work (Davis, 1993). Groups of four were used wherever possible, with a few groups of three required due to the number of students in the class. An advantage of opting for a group size of four it meant that even if one student dropped out, the dynamics of working in a group would still be preserved.

Groups of balanced mathematical ability were formed using a random allocation process, to avoid some groups being weaker or stronger. Heterogenous groups tend to perform better (Brickell, Porter,]Reynolds & Cosgrove, 1994). To create the groups I divided the class into quartiles (based on their performance in a diagnostic test) and then selected one group member from each quartile.

Student learning can be disadvantaged by creating groups where an individual feels isolated (Rosser, 1998). The second year I ran the initiative I also incorporated some gender balancing, so that no female was on their own in a group (there is a lower percentage of females in the class which means without the balancing it would be relatively common that a group would consist of one female and three males). This meant teams were either all male or an equal split of gender. It was interesting to hear a student comment that it was the first time in her university career that she had ever had another female in her group, so this balancing was obviously appreciated. Further balancing of other minority groups may also be desirable, particularly in light of the group which disintegrated the second year I rang the project. Language barriers and cultural differences were significant factors in the group's dysfunction, with one person having a different first language from the others.

To allow the groups to function well together, I focussed an entire two hour tutorial session on team building exercises. Students were introduced to their fellow team members and worked through a number of ice breaker exercises designed to get them talking and discovering each other's strengths. During the session they also worked on several maths problems together and we discussed how the project would be marked.

One problem with the team building session was how to deal with people who were absent. If one person was absent it wasn't a serious problem but if two or more members of a group were missing from the tutorial this caused issues. In several cases I ended up getting individuals and pairs to tag onto another group, so they could still be involved with group discussions and dynamics. An unintended consequence of this approach was the students developed a bond with the students in their temporarily adopted group and then wanted to stay together. Unfortunately I wasn't able to permit as the groups had already been predetermined. One solution to this problem would be to make attendance at the tutorial worth marks (which would likely result in close to 100% attendance). Alternatively I could dynamically assign people to groups, if I had their diagnostic test results on hand.

I purposefully choose a question topic that was open ended yet could be answered quantitatively. I also wanted a question that would be easy to understand yet relatively challenging to answer accurately. I had stressed the importance of assumptions in modelling and also the iterative nature of modelling, whereby you progressively improve your answer. A good question would allow both simple and sophisticated modelling approaches, depending on the assumptions the team made (and what factors they built into their model). Ideally every team would be able to come up with a rough approximation to the answer and then improve on that over the course of the day. Through running the project and NZ's Next Top Engineering Scientist I am developing a bank of questions of this nature. Some are extremely simple to phrase:

- *How far do you have to dive underwater to be safe from gunfire?*
- *How much fuel is required for a manned mission to Mars?*
- *If a severe Tsunami warning was issued, how long would it take to evacuate the 13,000 people who live on Te Atatu Peninsula?*

Others require a little more background information to be given as part of the question:

- *In the Disney Pixar movie "Up", 78 year old Carl Fredricksen attaches a large number of balloons to his house and flies it away. How far would a house lifted by balloons travel before landing?*
- *The aim of Project Loon is to ensure everyone on the planet has access to the internet, by creating a balloon-powered network. How many balloons would be required to provide balloon-powered internet coverage to all of New Zealand?*
- *Felix Baumgartner plans on breaking the world record for high altitude skydiving. He will make his jump from a capsule suspended beneath a balloon, at the edge of space. After Felix has landed, a remote triggering system will release the capsule from the balloon. In the event that electronic tracking is unavailable, what size search area is required in order to retrieve the capsule?*

In all cases there is a measurable quantity involved, so they will need to find a numerical answer and more importantly justify it.

Each group submitted a single report, after working on the problem for the duration of the day. Although the question wasn't revealed until the start of the day, students were provided with a rubric on how the written report would be marked, well in advance of the project day (please contact me if you would like a copy of this rubric). Students were also told that by default all members of a group would receive the same mark. This was done to incentivise working well together however I felt it was important to have a safety net in place in case of "free-riding" where individuals received a grade for the project without making an appropriate contribution. I adopted a peer review process, as peer review has been shown to be very effective in eliminating any issues related to free-riding (Dyrud, 2001). Each student ranked their own contribution and the contributions of their peers on a four point scale, which was submitted to me. The other members of the group did not get to see these rankings but students were informed that if the peer review indicated that the contributions were not well balanced I would contact the group members to discuss how best to mark their individual contributions. The first year I ran the group project there were no problems. In the second year I had two groups with issues, both of which were resolved after discussion with the students. In one case a student had refused to participate, so they received a mark of zero. In the other case the group had become very dysfunctional and had failed to submit work of a decent quality, so I replaced the project mark for all group members with an increased weighting on the exam. Another possible approach if contributions are uneven is to assign individual marks based on a divided mark approach where students receive a share of the mark based on their share of the work. Students perceive this method as “a fair way of arriving at grades for group work” (Maiden & Perry, 2011).

Was it a Success?

It is pleasing to note that the two times I have run the day-long group project correlate to the only two instances where every single student has passed the course (typically I expect a failure rate of around ten percent). Of course correlation doesn’t imply causation but the summer school version of Mathematical Modelling 1 has been relatively stable in terms of both content and teaching staff. The day-long group project has been the most significant new initiative so perhaps it is a contributing factor in getting everybody through.

As Summer-school 2012 marked the introduction of the day-long group modelling project into Mathematical Modelling 1, I was extremely interested in what students would think of this form of assessment. I discussed the format of the project early on and then recorded student feedback at three points via short anonymous questionnaires (at the start of the course, just after the project and at the end of the course). Unsurprisingly students’ perceptions changed over time. In the first week of class they had a very positive outlook (perhaps I had done a good job on “selling” the benefits of the project). The day after completing the project student perceptions of the benefits had dropped (although the outlook was still generally positive). By the end of the course opinion had lifted slightly with the majority agreeing that “taking part in day-long group project day was a valuable learning experience” and that “the project helped increase my understanding of the mathematical modelling process”.

As a teacher I was extremely pleased with the general performance of the groups. With the exception of one dysfunctional group in 2013, students demonstrated excellent mathematical modelling skills. The quality of the submissions were relatively high, especially given that students only had a standard working day in which to produce a report. The reports demonstrated that the teams had successfully applied their modelling skills to a larger scale problem, which was the outcome I had hoped for. Interestingly students were generally unhappy with their performance, with several people commenting they could have achieved more if they had been given more time. I think the time constraint is a valuable part of the experience though, as it adds an interesting element to the assessment. The project isn’t just about finding the best answer, it is about finding the best answer subject to the time constraints.

Most students were happy with the mark their group received but when surveyed immediately after the project day fewer than half of the students were happy with the quality of their work that they had submitted. This may reveal students’ inability to accurately self-assess their performance in a modelling project, despite being provided with a fairly detailed rubric. My hunch is that students were probably more concerned with getting “the right answer” as opposed to what I was assessing, which was how well they had applied and understood the mathematical modelling process.

Given that the students generally performed very well and enjoyed working as part of a team it was a little disheartening that fewer than half the students surveyed wanted to do similar assessments in the future. It would be interesting to delve further into why they didn't want to repeat the experience. Overall student feedback was positive enough that I have decided to continue the initiative. For a more detailed analysis of student perceptions, see my paper on *Student Perceptions of a Day-Long Mathematical Modelling Group Project* (Bier, 2012).

References

Bier, P. (2012). Student perceptions of a day-long mathematical modelling group project. *Proceedings of the Australasian Association of Engineering Education 2012 Conference, Melbourne, Australia,* retrievable from: http://www.aaee.com.au/conferences/2012/documents/abstracts/aaee2012-submission-182.pdf

Borri, C., & Maffioli Firenze, F. (2007). *Re-engineering Engineering Education in Europe*. Firenze University Press.

Brickell, J. L., Porter, D. B., Reynolds, M. F., & Cosgrove, R. D. (1994) Assigning students to groups for engineering design projects: A comparison of five methods. *Journal of Engineering Education*, *83*(3), 259–262.

Davis, B. D. (1993). *Collaborative learning: Group work and Study teams. Tools for teaching*. San Francisco: Jossey-Bass.

Dyrud, M. A. (2001). Group projects and peer review. Business Communications Quarterly, *64*(4), 106–12.

Rosser, S. V. (1998). Group work in science, engineering and mathematics: Consequences of ignoring gender and race. *College Teaching*, *46*(3), 82-88.

Maiden, B. & Perry, B. (2011). Dealing with free-riders in assessed group work: results from a study at a UK university. *Assessment & Evaluation in Higher Education*, *36*(4), 451-464.

Author
Peter Bier, the University of Auckland, New Zealand.
Email: p.bier@auckland.ac.nz

A Report on the Use of STACK in Mathematics at Birmingham 2012–2013

C. J. Sangwin & D. F. M. Hermans
School of Mathematics, University of Birmingham, United Kingdom.

Background

Valid assessment is the cornerstone of effective teaching. Automatic assessment is commonly associated with multiple choice questions (MCQ) or similar provided response question types. Such question types are referred to as *objective* because the outcome is independent of any bias by the assessor. MCQs have been criticized for many years: Hoffmann (1962) claims they *"favour the nimble-witted, quick-reading candidates who form fast superficial judgements"* and *"penalize the student who has depth, subtlety and critical acumen"*. Authors such as (Hassmen & Hunt, 1994) and (Leder, Rowley & Brew, 1999) provide evidence that the MCQ format itself has inherent gender bias.

For mathematics MCQs are particularly problematic as the relative difficulty of a reversible process, e.g. integration compared to differentiation, is markedly altered in different directions. As (Sangwin, 2013) says

> The strategic student does not answer the question as set, but checks each answer in reverse. Indeed, it might be argued that it is not just the strategic, but the *sensible* student, with an understanding of the relative difficulties of these processes, who takes this approach. This distortion subverts the intention of the teacher in setting the question, so that we are not assessing the skill we wish to assess. Hence the question is *invalid*.

While the term 'objective test' is often taken to be synonymous with MCQs there are many situations, particularly in mathematics, where the properties of an answer provided by a student can be established objectively and automatically. For these reasons, where practical in mathematics, automatic assessments which accept a mathematical expression from the student are to be strongly preferred.

Motivated by the need to assess answers to students' work, the STACK online assessment system uses computer algebra to support the assessment process. It also

- generates random versions of questions in a structured mathematical way;
- accepts answers from students which contain mathematical content, rather than MCQs;
- establishes the mathematical properties of those answers;
- generates outcomes which fulfil the purposes of formative, summative and evaluative assessment;
- stores data on all attempts at one question, or by one student, for analysis by the teacher.

In particular STACK uses the computer algebra system Maxima to support the mathematical processes, and the Moodle learning environment to provide a context in which the activity takes place (see Wild, 2009). A demonstration server is available at http://stack.bham.ac.uk/moodle.

Given the student's answer consists of a mathematical expression, questions are raised about how students are to actually enter their answer into a machine (see Sangwin &

Sangwin, C. J. & Hermans, D. F. M. (2013). A report on the use of STACK in mathematics at Birmingham 2012–2013. In B. Barton, G. Oates, & M.O.J. Thomas (Eds.), *Community for Undergraduate Learning in the Mathematical Sciences (CULMS) Newsletter*, ***8***, 16-30.

Ramsden, 2007; Sangwin, 2012). This is an important issue, and we believe that first year mathematics undergraduates should learn how to type simple expressions into a machine accurately using a relatively strict syntax.

The prototype test is to establish that the student's answer is (i) algebraically *equivalent* to the correct answer and (ii) is in the appropriate *form*, (e.g. factored). That said, the use of computer algebra is a long way from string matching, or the use of regular expressions. It is the relevant properties of expressions which matters and answers are often non-unique. Where the student's answer does not satisfy all the properties, the teacher is able to encode feedback. Potentially this is specific to the answer and directly related to possible improvement on the task. This is precisely the kind of feedback which research such as (Kluger & De Nisi, 1996) have suggested is most effective. Unusually for CAA, STACK may include and display results of computer algebra calculations within such feedback which can be as detailed as appropriate to the situation. This is a particular distinguishing feature of STACK.

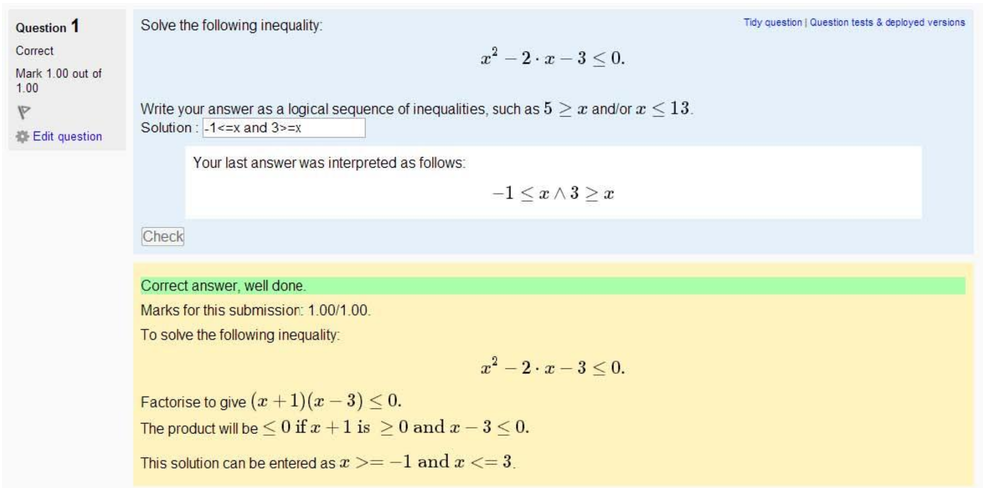

Figure 1: A STACK from Assessed Quiz #3: Inequalities, question 1

Online assessment using computer algebra was introduced to Birmingham by Dr Hermans in 1999. The current software, STACK, was designed and developed by Dr Sangwin at the University of Birmingham starting in 2004. During 2011-2012, version 3 was written in collaboration with Tim Hunt at the Open University, with contributions from Matti Harjula at Alto University, Helsinki and Matti Pauna at the University of Helsinki. More details of the goals and design are given in Sangwin (2013).

Use in Birmingham Mathematics 2012–2013

In this section we record the way we have used STACK in assessment of first year mathematics students at the University of Birmingham during the 2012–2013 session. STACK was used in the first-year core module "Calculus and Algebra". This is a 40 credit module (where 120 credits is a full year load), split between two eleven week semesters. STACK quizzes were use in three ways; (1) open access practice, (2) to support computer labs and (3) as formal assessed coursework. The cohort consisted of 231 students (plus two "test students", i.e. teachers with student accounts for testing purposes).

We believe learning to write simple expressions in a formal syntax is an indispensable skill for undergraduate mathematics students, but one which we have a responsibility to teach. Hence, an explicit *"syntax practice"* quiz is available to help students learn how to enter their answer into STACK. The quiz consists of "questions" such as *"The constant* π *is entered as* pi. *Type this in here."* There were 216 attempts at this quiz.

Open access practice

Nine open access practice quizzes, giving a comprehensive coverage of formal symbolic differentiation, were made available in semester one. The number of attempts at each quiz is shown below. The low level of take up is indicative of our perception that students do not engage with open access Nine open access practice quizzes, giving a comprehensive coverage of formal symbolic differentiation, were made available in semester one. The number of attempts at each quiz is shown in Table 1. The low level of take up is indicative of our perception that students do not engage with open access

Table 1: Number of attempts at the nine practice quizzes.

Quiz	# attempts
1. Derivatives of basic functions	52
2. Linearity in differentiation	16
3. Higher derivatives	13
4. The product rule for differentiation	7
5. The quotient rule for differentiation	7
6. The chain rule for differentiation	5
7. Implicit differentiation	9
8. Parametric differentiation	6
9. Miscellaneous differentiation	6

Computer labs

Students are required to learn how to use the Maple computer algebra system (Heck, 2003). STACK quizzes support this activity, by asking students to solve typical problems using the Maple commands introduced in the associated Maple worksheet. This aims to help students to engage with resources and turn a session, which so readily could be a passive reading through, into an interactive investigation. Resource limitations did not allow us to provide adequate staff or postgraduate supervision during these computer lab sessions, so they were offered purely as a formative exercise.

There were 10 computer labs, but only those shown in Table 2 had accompanying STACK quizzes. It can clearly be seen from this data that a significant group of students did not engage with the computer lab quizzes in the second semester, perhaps because the assessed work did not contribute to the grade for this module.

Table 2: Engagement in the computer lab quizzes.

Quiz	# questions	# attempts	# students completing quiz	% average mark
1.	4	299	213	86.5%
3.	6	224	204	84.5%
4.	6	10	8	71.2%
5.	10	208	175	67.8%
6.	10	198	180	37.6%
8.	10	155	151	59.2%

9.	5	140	133	66.3%
10.	5	108	107	74.7%

Formal assessed coursework

10% of the formal assessment for this module was given for results to nineteen on-line quizzes. Nine occurred in semester 1 covering the following topics: sets, functions, inequalities, limits, derivatives, more functions, matrices, matrix inverse and limits and series. In semester 2 the topics included complex numbers (2 quizzes), symbolic integration (3 quizzes), matrices and determinants, first and second order ODEs, matrix inverse, vector spaces and conics.

Quizzes enable students to demonstrate a *minimum competency threshold* on indispensable procedural skills. Hence, questions were marked either 0 or 1 and the final mark was the *product* of all the individual question marks. I.e. students were required to score 100% on *all* questions to gain the 10% towards the module mark. Students could re-take quizzes as often as they wished, and the number of attempts at each quiz is show in Table 3.

The cohort consisted of 231 students. Of these 6 students completed nothing, and 221 students completed the formal assessed coursework satisfactorily. Therefore only 4 students making a sensible attempt failed to get the 10% of the credit for this module.

Table 3: Response to the Semester 1 assessment quizzes.

Semester 1 Quiz	# questions	# attempts	# completed quizzes
1. Sets	7	288	232
2. Functions	6	301	230
3. Inequalities	6	263	231
4. Limits	5	245	231
5. Derivatives	6	282	230
6. More Functions	6	244	230
7. Matrices	5	236	230
8. Matrix Inverse	4	234	230
9. Limits & Series	5	236	229

These numbers (and the equivalent results for Semester 2 in Table 4) seem to indicate that most students got to grips with Moodle–STACK very quickly. In theory, each student needs only one attempt at a quiz, as each question allows for an unlimited number of attempts. The need to re-attempt a quiz arises when students inadvertently ‘finish’ a quiz, i.e. indicate they have finished with a quiz, rather than leave the quiz open till all questions have been successfully answered. Where this occurred, the numbers above suggest its incidence dropped off rapidly. In addition, students may reattempt quizzes just for further practice. Only the quiz with the highest score was counted.

The quizzes on material taught during the first semester were due in on 13th January 2013 at 23:50 hours. The quizzes on material taught during the second semester were made available between 22– 29 March, and had a deadline of 29th April. Figure 2 shows the gradual completion of the second semester quizzes by the deadline of 29th April. The final completion rates of 95% represent 221 out of 232 students. As can be seen from this data the majority of students worked consistently during the Easter vacation to complete the online assessed work before the deadline.

Table 4: Response to the Semester 2 assessment quizzes.

Semester 2 Quiz	# questions	# attempts	# completed quizzes
1. Complex Numbers 1	5	239	225
2. Integration 1	4	234	223
3. Integration 2	5	238	223
4. Integration 3	2	229	224
5. Matrices & Determinants	8	226	222
6. First Order ODE’s	4	233	221
7. Second Order ODE’s	7	225	222
8. Vector Spaces	10	225	221
9. Conics	6	228	221
10. Complex Numbers 2	7	232	221

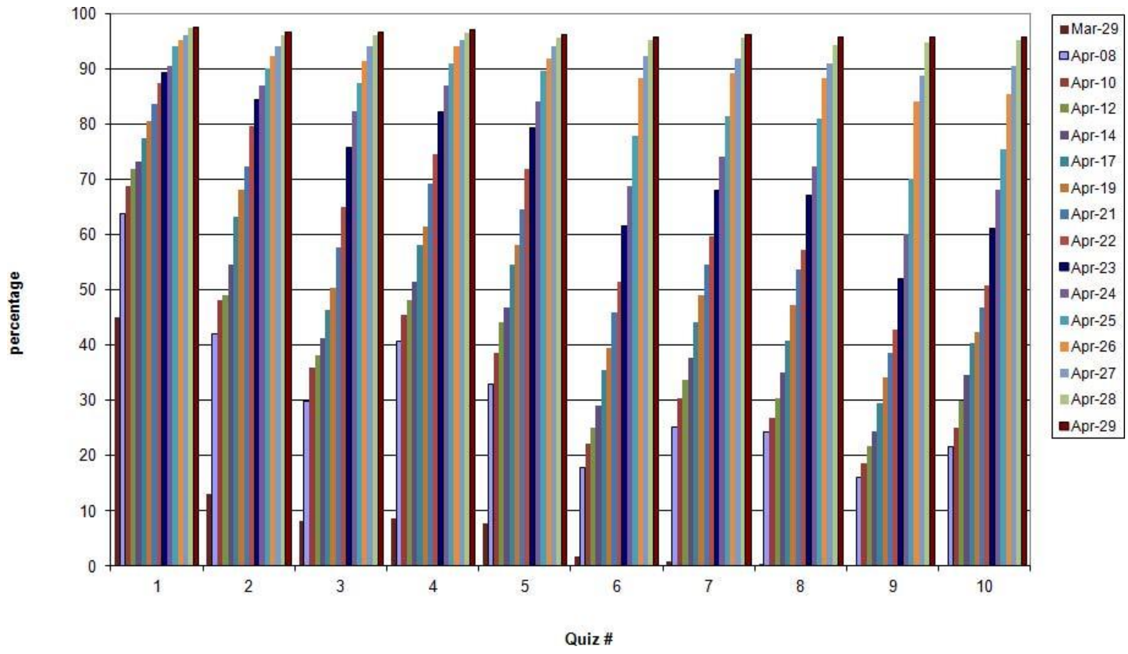

Figure 2: Completion rate of Semester 2 quizzes during the final days of April 2013.

Authoring Questions

For all the quizzes that were envisaged, questions existed in the legacy AiM format (see Klai, Kolokolnikov &Van den Bergh, 2000; Strickland, 2002). An internal project (From AiM to STACK, CLAD, CLP025) provided funds for a Learning and Teaching Consultant who converted these questions into STACK format. This was helped by a semi-automatic conversion tool that would populate various fields in the STACK question format with information from the AiM question source. Further editing was required, mainly in converting the question setup from Maple to Maxima and to establish the question response trees with appropriate feedback. Once the consultant had converted questions, they were then quality controlled by the course leader, Dr D. F. M. Hermans, before being assigned to a quiz. The conversion of the questions from AiM to STACK also created an opportunity to change the delivery style. Over the past decade or so, assessed quizzes referred to periods of delivery, typically a quiz every week in which both lessons in algebra and calculus took place. Hence each AiM quiz typically contained questions over at least two topic areas. Apart from the lab related questions, which continue to be grouped by computer lab setting, the new STACK questions were grouped in quizzes by topic, which makes it easier to analyse strengths and weaknesses of a student's performance across topics.

In the first semester, time pressures, and a need to learn more about aspects of Maxima, resulted in a small number of existing AiM questions not being converted (11 out of 63) and a further 5 questions not finishing their quality control in time. However, all 58 AiM questions for the second semester assessments were successfully converted and deployed, as were all the 56 AiM questions relating to Computer Laboratory work. One multiple choice question was converted to a Moodle multiple choice type, which did mean that data could not be randomized. There is currently no STACK version of the multiple choice questions which allows for randomization of question data.

In addition, work is on-going to author derivative questions. These are questions that are similar to an existing question but have their randomization restricted to cater for, e.g. a particular case. For example, an existing question to solve a quadratic equation can lead to three derivative questions, one each for the case of a quadratic equation with two real roots, a double real root, or a pair of complex conjugate roots.

One very helpful feature in STACK is the ability to check each question instance automatically. Inspired by the practice in software engineering of unit testing, one can set up a number of different answers (some correct, some incorrect) with the associated intended response. These can be automatically checked when randomizing the question data, to ensure that the feedback structure is correct for every single question instance.

A new aspect of the Moodle–STACK integration is the ability to tailor the feedback to the context in which the question is posed. Moodle allows for several 'question behaviours' and for each one, detailed settings can be customised. These include e.g. marks and the correct answer; specific and general feedback. STACK fully uses these options and maps the specific feedback to the feedback it generates in response to the analysis of student answers. General feedback can being used to display a fully correct worked out answer, but using the students' particular data, so it is tailored to each individual student. This integration with Moodle management of question behaviour allows for a flexible tailoring of feedback to the type of context you are planning, e.g. exam style quizzes (summative) or practice quizzes (formative).

Question scopes are essentially limited by what Maxima can do, both in the setup of questions and in the analysis of answers. For example, Maxima (currently) has no way to enter real sets using interval notation. We could not find an elegant way to check if a matrix given is in echelon form and still represents an equivalent system of equations. This is an inherent boundary to any assessment system that combines a computer algebra package with an assessment tool. Another aspect of question authoring that requires careful analysis is to pick the best structure for the response tree. What answers, correct or typically wrong, do we want to identify and provide feedback on, and how can this best be achieved? This requires careful thought by the teacher, and would be the case regardless of the CAA system used.

Academic Support

Students were able to seek support with any mathematics issues before or after lectures, during weekly exercise classes and tutorials as well as during scheduled student contact time with the lecturers involved. In addition, and typically when a deadline was approaching, they also used email to contact the lecturers with any issues. The course leader was the main target for any questions. In semester one, he received 16 queries from 14 different students, in semester two, this increased to 25 queries from 19 students, out of a total cohort of 231 students. Of these 41 queries, 9 were seeking confirmation that marks had been recorded or queries on how they contributed to the module mark. Two queries related to an issue with the registration status of students, and 30 related to a mathematical issue, either not understanding what the question was asking for, or answers that the students believed were right, but were (correctly) marked wrong. This feedback is useful in understanding the issues students face with the questions posed. In some cases, this can lead to a better formulation of questions, or additional feedback being provided by editing the question.

In two cases, answers that were equivalent to the correct answer were not marked right. The first is a solution of an indefinite integral:

$$\int \frac{1}{x} \ln\left(\frac{9}{x}\right) dx$$

which can be expressed as:

$$-\frac{1}{2}\ln^2\left(\frac{9}{|x|}\right) + c = -\frac{1}{2}\ln^2\left(\frac{1}{|x|}\right) + ln9ln|x| + c.$$

One can identify that both expressions have a derivative equal to the integrand, but so would:

$$-\frac{1}{2}\ln^2\left(\frac{9}{x}\right) + c$$

which we insisted on marking as incorrect. This type of question will require an identification in the student answer of the absolute value signs (or equivalent).

The second question was a first order differential equation:

$$\frac{dy}{dx} = 35 + 5x^2 - 7y^2 - x^2y^2$$

which leads to the direct integration:

$$\int \frac{1}{5 - y^2} dy = \int (7 + x^2)\, dx.$$

The left hand side integral can either be solved as:

$$\frac{1}{\sqrt{5}} \tanh^{-1}\left(\frac{y}{\sqrt{5}}\right),$$

or:

$$\frac{1}{2\sqrt{5}} \ln\left(\frac{\sqrt{5} + y}{\sqrt{5} - y}\right).$$

This leads to two quite distinct expressions for y. Of course, both cases can be responded to by allowing all equivalent answers to be checked for in the response tree.

Issues like these are brought to the fore by student queries, but many more instances may occur which the student can easily overcome themselves. Equally, there may be typical wrong answers that are not currently identified which could be captured and accounted for in the response tree. STACK enables teachers to view all the different student answer attempts, grouped by random question version, for further analysis. These features need further work to provide a range of basic statistics, but all data is available to the teacher for subsequent use in robustness checking, quality control and improving formative feedback.

Students' Reactions

Students taking the course were invited to complete an online survey to gauge their reactions to STACK. This consisted of nine questions to which students responded on a Likert[1] scale.

1. I generally complete all the coursework.
2. The syntax for entry of answers was easy to learn initially.
3. I am now confident in entering my answers into STACK.
4. The feedback given by STACK helps me improve my performance on the tasks.
5. Doing STACK questions (labs & exercises) has helped me with my understanding of the module.
6. Doing STACK questions (labs & exercises) meant that I engaged with tackling problems which I would not otherwise have done.
7. The STACK questions in the computer labs have helped me to familiarise myself better with the content of the Maple worksheet.
8. Overall, STACK quizzes provide a quick and useful feedback mechanism for the module.
9. I would be happy to have STACK mark the final exam for the module.

The last question was an open ended text response to:

> Please, add any comments or suggestions you may have, or, any more detailed explanation to your answers to the questions above that you may wish to share.

145 replies were received from 231 registered students giving a response rate of 62%. The quantitative results to individual questions are shown in Table 5. Since 96% of respondents agree or strongly agree that they "complete all the coursework" we have not undertaken an analysis of other responses correlated with this question.

It is clear from Table 5 that the majority of students are satisfied with STACK. The syntax was not a serious barrier initially, (or perhaps the instructions we gave were effective) and only 6 students disagree or strongly disagree with Q3, indicating that they are still not confident in entering their answers. The majority of students agree or strongly agree that the STACK questions helped their understanding of the material and improved their performance both on these tasks and others. Overall 66% of student agreed STACK quizzes provide a quick and useful feedback mechanism for the module. In the more specific questions, 2–7, the majority of students agreed or agreed strongly, except in the issue of feedback where only 49% took this view. We will comment on feedback later.

Excluding Q9, insignificant numbers of students disagreed strongly, but 47 individuals disagreed with something. One individual strongly disagreed with Q3, Q5. and Q6.

[1] i.e. strongly agree, agree, neutral, disagree, or strongly disagree.

Another individual disagreed strongly with everything except Q7. (where they only disagreed) and Q1. (where they agreed strongly that they completed all the work). His/her comments to Q10. were as follows:

> The method of input to Stack for the quizzes is extremely cumbersome. For an example, you have to be much more explicit when entering answers than for a program like Wolfram Alpha. The questions which arise on the quizzes are mostly just what we would have done in our examples worksheets, but in some cases are much easier. I don't find the quizzes useful, but it's an easy way to get 10%; of the course so who can complain?

Table 5: Survey responses for students' reactions to STACK.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1. complete coursework	96 (66%)	43 (30%)	3 (2%)	3 (2%)	0 (0%)
2. easy initial syntax	25 (18%)	86 (63%)	20 (14%)	5 (4%)	1 (1%)
3. now confident syntax	32 (23%)	76 (55%)	24 (18%)	4 (3%)	2 (1%)
4. feedback improved performance	11 (8%)	62 (43%)	46 (32%)	20 (14%)	5 (3%)
5. helped understanding	16 (11%)	72 (52%)	36 (26%)	13 (9%)	3 (2%)
6. helped unfamiliar problems	20 (14%)	76 (55%)	29 (21%)	13 (9%)	2 (1%)
7. helped with worksheets	15 (11%)	90 (64%)	20 (14%)	12 (9%)	3 (2%)
8. Overall useful	21 (15%)	70 (51%)	26 (19%)	17 (12%)	5 (3%)
9. Happy with STACK exam	11 (7%)	22 (15%)	50 (34%)	47 (32%)	18 (12%)

There appear to only be 2 students in 145 survey respondents who are significantly disgruntled with STACK, whilst the vast majority view it as a helpful contribution to their education in this module without significant barriers to use.

In general students were much less happy with the suggestion that STACK might be used to assess an exam for the module. Less than one quarter agreed, and nearly half disagree or strongly disagree with this suggestion. As we shall see, this issue is evident in the open-ended responses to the last question.

There were 67 non-empty comments to question 10. These have been grouped in broad themes as follows:

The largest single group of 15 were simply unqualified positive comments, e.g. *"Enjoyed them, thought they really helped my understanding"*. A further 8 were positive but qualified, e.g:

> STACK is good for questions with low amounts of working of any complexity, but less useful for problems with longer and more complex working.

and

> The STACK are good for reassurance of whether a topic has been understood. The syntax can be frustrating, and little feedback is given.

These significantly outweighed 7 negative comments such as *"I still much prefer working on paper and having a lecturer or PHD student mark my work than a computer. For example, [...]"* and *"STACK is cruel and evil."* One student expressed a strategic view of learning, which doesn't really relate to STACK specifically:

> I do not really find stack quizzes useful and would prefer to only have to do the ones that count towards the module. i.e. I do not like the unassessed maple quizzes and exercises.

This was counterbalanced by other students who would prefer assessments which do not contribute, e.g. *"Keep them un-assessed."* Ten students commented on the difficulty of the syntax, e.g:

> I find the STACK quizzes useful as they make me do more revision and look over things so I can answer the questions. However sometimes the answer is right I'm just not sure how to input it and this can be confusing.

There were a number of comments which might be interpreted as relating to difficulties with *syntax* or might be interpreted as a difficulty in understanding the correct *form of the answer*:

> STACK is so picky with formatting! The right answer can be marked wrong for petty reasons!

and

> As good as it is providing instant feedback, it can be annoying at times. I remember this question: $\sqrt{(x^2 - 9)}$ what is the domain? *x*>=3 and *x*<=-3 was wrong, I had to use "*or*" instead of "*and*", and it took me longer to work that out than it did to all the other questions.

Note, what the student interprets as syntactic nit-picking is actually a matter of correctness. In addition, as we gain experience of what students actually type in we can check for the occurrence of particular incorrect answers, such as this, and provide feedback tailored to the error/misconception which is likely to have given rise to the response. Other, more specific issues, may well require further work on the system, e.g:

> Sometimes STACK seems to have issues with answers that are essentially correct -(once I multiplied 2 square roots together i.e. sqrt((x-3)*(x-5)) and it said my answer was incorrect but then when I did sqrt(x-3)*sqrt(x-5) that was correct. It wasted time because I thought my calculation must have been wrong and was puzzled for a long time. There was also a problem with using (1/cos(x)) instead of sec(x) for something, which again can be misleading when it comes out incorrect as you don't know whether your maths was incorrect or they are just looking for a slightly altered expression.

What does *"slightly altered expression"* mean? The expressions $\sqrt{(x-3)}\sqrt{(x-5)}$ and $\sqrt{(x-3)(x-5)}$ have different domains, so potentially define rather different functions. This is quite different from a synonym of sec(x) for the reciprocal of cos(x), although it is not clear if this student appreciates this. It is an issue which is often glossed over in elementary teaching. When, and why, does the written format of an answer matter? Better availability of both feedback in these situations, and face to face help, might alleviate these frustrations. To the authors, the "pickiness" of the system is one of its strengths. Students need to learn to be unambiguous, and to check their own work. If we take this view we do have a responsibility to make sure the system actually works robustly.

Eight students asked for more specific feedback. For example:

> It would make things much easier if it clearly stated in which form you should give your answers, and clearly stated when and how you go wrong.

One student acknowledged *"Knowing if you were right or wrong instantly is useful."* Writing STACK questions requires an investment in effort. The basic assessment design needs to be completed, with attention to the format of the assessment. This design is required for all formats, including paper, multiple choice etc. Then random versions, with corresponding full worked solutions, need to be written. The highest priority in year one is that the assessment algorithms robustly assess correct answers as correct, and incorrect answers as incorrect, and that where the form of the answer is important (e.g. factored, partial fractions) this is implemented correctly. Implementing hints, suggesting methodological or conceptual mistakes consistent with the actual answers provided requires significant extra effort. Actually, it is much more efficient to write these in year 2, when you have trialled materials with a full cohort. So, while the teacher can (and does) anticipate some incorrect responses, developing high quality STACK materials requires at least two full cycles, and a commitment to review and revision. We should also add that the quizzes occur as part of a coherent teaching scheme, and that tutors were available both face to face (in weekly tutorial meetings) and by email to consult. The email queries in particular also form a useful source from which to revise the materials for use in year 2 and beyond.

A number of student commented on the links between the Maple worksheets and the STACK quizzes.

> It gets us used to Maple, it was a more gentle way in than just going straight to the maple assessed exercises. I had never used Maple or anything close, so that was good for me. It also is quite unforgiving, if you accidently put the wrong number, you have to redo the whole quiz, it would be nice if you got 2 chances, just in case that does happen.

Ten of the comments related to using STACK for assessments of exams. Many students expressed sophisticated views of the strengths of STACK for this purpose.

> I think Stack is a useful tool and the Lab worksheets helped familiarise me with Maple. I don’t think that I would like Stack to mark the final exam because I don’t think it can replace a human marking the paper -I don’t think it is smart enough yet and you would be restricted by the kind of questions that you could set so that you can answer in a syntax that Stack can pick up.

and

> Regarding question 9: I would be happy to have STACK marking the assessed work, however I believe a human judgment is needed with some more lengthy questions so I would prefer for the marking to be kept as it is.

From another individual:

> The reason why I don’t want stack to mark the exams is because there are loads of different methods people may take in answering one question and if someone has done a different method stack might not recognise it and then they could lose marks for no reason.

Only one student was positive, but cautious.

> I think part of the final mark should be STACK based, but the final exam would be very difficult to mark entirely on STACK, as different solutions are possible, and we deserve credit for methods as everyone makes small arithmetical errors from time to time.

There were no comments relating to difficulties of accessing the server, or “lost work” or other infrastructure related issues.

Server Specification and Load

http://stack.bham.ac.uk/moodle is designed to provide a demonstration server for the STACK project and external users are welcome to register themselves on this server to evaluate STACK exercises. Staff and students at the University of Birmingham gain access to STACK quizzes through the IVLE (WebCT) using a bespoke mechanism which identifies them, authenticates their request to connect and establishes their credentials on the course (i.e. staff or student).

We are often asked "how big is your server?", so for the record the production server has 16 Intel Xeon CPU E5520 processors running at 2.27GHz. We used Maxima 5.28.0, through CLISP, running the direct connection in optimized mode, with the CAS cache. Without the CAS cache, running 505 answer-test unit tests took 87.7 seconds. With a mature cache, this took 10.9 seconds. We therefore estimate each CAS call costs 0.15 seconds, a significant overhead even on a fast server. Plots appear to take longer and since these require numerous floating point calculations this is to be expected.

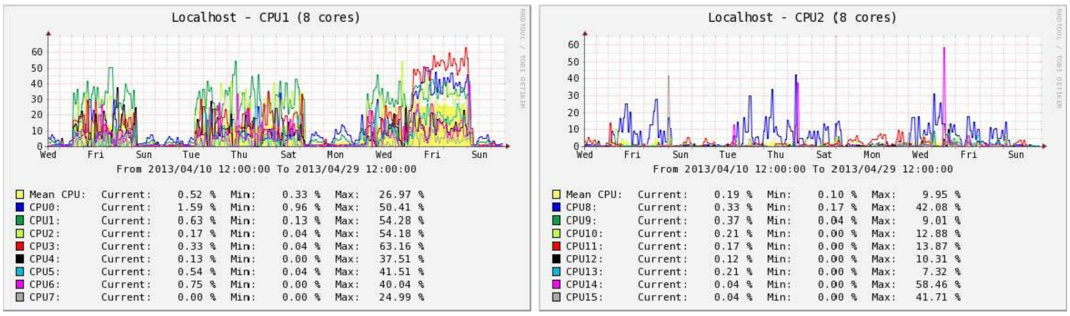

Figure 3: Load on the live server during the crucial end of session peak.

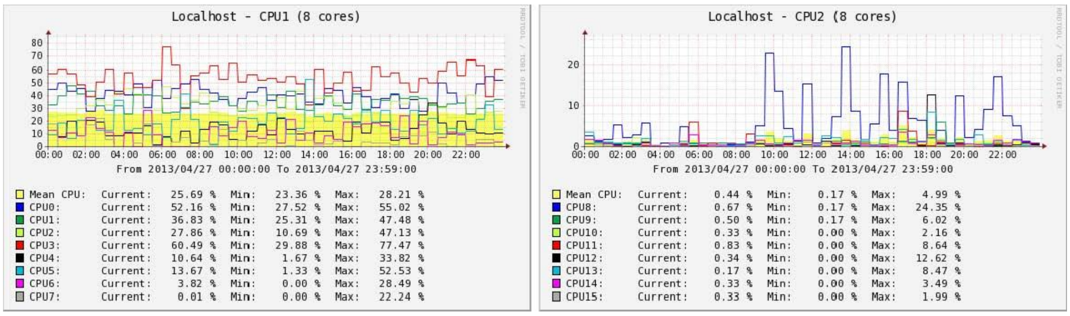

Figure 4: Load on the live server on April 27th 2013.

All assessed quizzes set during the second semester had a deadline of 29 April 2013, 12:59am. The graphs in Figure 3 show the processor load during the period of 10 April 2013, 12:00 until 29 April 12:00. This includes Saturday 27th April, perhaps the day with the highest server load, which is shown in Figure 4. On that day the server recorded 3940 submit requests by 44 individual students, not including the teacher or a guest user. A histogram of the times of access are shown in Figure 5. Notice the uniformity of load during the day. Furthermore, students were working on a variety of topics during that day as shown in Table 6.

11 students made fewer than 20 interactions on the server. 10 made between 20 and 100 interactions, and 14 students made more than this. Students making multiple interactions are attempting STACK questions, i.e. not just login and navigation requests.

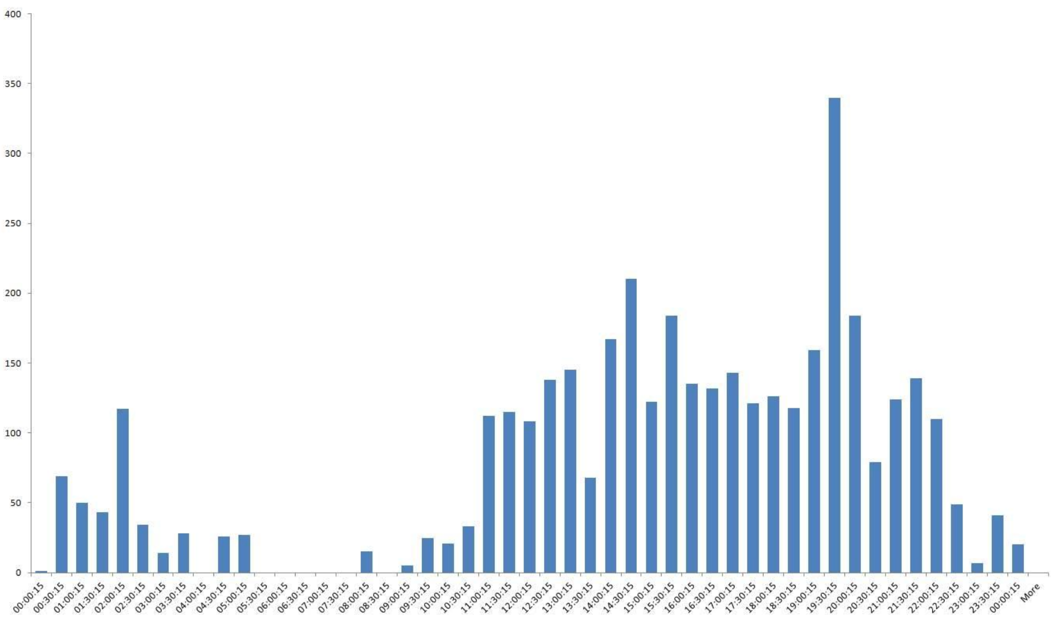

Figure 5: Access times on the live server on April 27th 2013.

Table 6: The number of attempts at, or reviews of, questions in the assessed quizzes during April 27th.

	Semester 1	
1.	Sets	17
2.	Functions	7
3.	Inequalities	1
4.	Limits	0
5.	Derivatives	0
6.	More functions	0
7.	Matrices	0
8.	Matrix inverses	1
9.	Limits and Series	0
	Semester 2	
1	Complex Numbers 1	79
2.	Integration 1	130
3.	Integration 2	186
4.	Integration 3	89
5.	Matrices and Determinants	167
6.	First Order ODE’s	266

Future Directions

Over the past session, we successfully changed completely from AiM to STACK as the delivery mechanism for on-line mathematics quizzes. Online quizzes will continue to be available to students as a practice aid. However, we are planning to increase our use of

STACK quizzes as part of summative assessment. In particular we are looking at developing “competency tests” in STACK. These would test the ability and fluency of solving standard problems, e.g. how many derivatives can you calculate correctly in a given amount of time? It is envisaged that completing such competency tests at a sufficient high threshold would become a requirement for passing the module. In this way, a traditional written examination paper can be focussed on methodology, proofs and more conceptual questions.

The cycle has highlighted areas for future technical development, particularly for inequalities and intervals of the real line. We also intend to update and improve feedback for a number of questions in the light of students’ answers during this cycle.

Conclusion

The server infrastructure was entirely reliable throughout this testing period. Very few “bugs” in the STACK software were encountered, although experience with students enabled a number of significant improvements to the interface, especially for real inequalities, to be made during the year. The majority of students are satisfied with STACK, eventually are confident in using the syntax and think it is a valuable *part* of their course. Students are less happy with the prospect of using STACK to automate examinations. It is also clear that students do not engage with open access materials which do not carry credit for the course.

The use of STACK based on-line assessment has allowed the School to create a provision of exercise opportunities with quality feedback to assist the students with their learning. This provision would not be possible in the labour intensive traditional homework with manual marking scenario and hence has added to the resources for our students. It has also enabled the School to accommodate individual learning styles of their students, with many really enjoying and benefitting from the opportunities offered.

Acknowledgements

We would like to acknowledge the Centre for Learning and Academic Development (CLAD) for supporting the translation of questions from AiM to STACK format (CLP025), and Mark Cooke and Michael Finney for supporting the server infrastructure within the School of Mathematics.

References

Euler, L. (2006). *Elements of Algebra*. Tarquin Publications.

Hassm´en, P. & Hunt, D.P. (1994). Human self-assessment in multiple choice. *Journal of Educational Measurement, 31*(2), 149–160.

Heck, A. (2003). *Introduction to Maple*. Springer-Verlag.

Hoffmann, B. (1962). *The tyranny of testing*. Crowell-Collier.

Klai, S., Kolokolnikov, T. & Van den Bergh, N. (2000). Using Maple and the web to grade mathematics tests. In *Proceedings of the International Workshop on Advanced Learning Technologies, Palmerston North, New Zealand, 4–6 December*, 2000.

Kluger, A. N. & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Leder, G. C., Rowley, G. & Brew, C. (1999). Gender differences in mathematics achievement: Here today and gone tomorrow?, In G. Kaiser, E. Luna & I. Huntley (Eds.), *International Comparisons in Mathematics Education*, **(**pp 213–224). Falmer Press.

Sangwin, C. J. (2012). The DragMath equation editor. *MSOR Network Connections, 12*(2), 5–8.

Sangwin, C. J. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press.
Sangwin, C. J., & P. Ramsden.(2007). Linear syntax for communicating elementary mathematics. *Journal of Symbolic Computation*, *42*(9), 902–934.
Strickland, N. (2002). Alice interactive mathematics. *MSOR Connections*, *2*(1), 27–30, 2002.
Wild, I. (2009). *Moodle 1.9 Math*. Packt Publishing.

Authors
Chris Sangwin, The University of Birmingham, United Kingdom.
Email: C.J.Sangwin@bham.ac.uk

Dirk Hermans, The University of Birmingham, United Kingdom.
Email: D.F.M.Hermans@bham.ac.uk

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau