

## An Experiment in Computer-Aided Assessment

---

Trevor Hawkes, Mathematics Institute

### BACKGROUND

Until three years ago, no courses in the Mathematics Undergraduate Programme at Warwick used continuous assessment bearing examination credit. With little incentive for regular work, students often failed to keep abreast of the material in their courses. Then, because mathematics unfolds as a hierarchy of concepts, these students quickly became unable to make sense of their lectures and grew disenchanted with university mathematics. The system tempted even the brighter students to postpone serious work until the revision period in the Summer Term and by this time it was too late for many to absorb and understand the huge backlog of undigested abstract mathematical ideas.

This situation has radically changed. Students are now required to submit regular weekly homework in each of the first- and second-year core courses. The scale of the change is well illustrated by the fact that over the past three years (1995-98) the number of pieces of assessed work for credit submitted by first- and second-year Mathematics students during the academic year has risen from zero to over 30,000. These assignments are marked by postgraduate supervisors and returned to students for discussion in supervision classes (of typically four students), usually within a week. (Lecturers moderate any disputes over the marking.)

The increased marking placed a considerable extra work load on the postgraduates who do the small-group undergraduate teaching. Moreover, additional payments for this marking have put an extra burden on the Department's teaching budget and diverted discretionary funds from other important priorities. An alternative approach was needed, not least to avoid a walk-out by the over-worked postgraduate teachers.

This is why I decided last year (1997-98) to experiment with the assessment in the first-year core course Foundations, which serves as a bridge between school and university mathematics and is taken by some 350 single-subject and joint-degree first-year mathematicians. Students continued to receive a weekly set of assignment questions, but instead of submitting answer sheets for marking, they were required to take a weekly multiple-choice test based on the assignment questions. Not only was the supervisors' marking load lightened but money was saved. The tests were machine-marked by Computing Services for only £10 a week. Set against a corresponding cost of £416 a week for postgraduate marking, this represented a saving of more than £4,000 over the term.

I was agreeably surprised by the effectiveness of the multiple-choice format in assessing students' knowledge and understanding of the basic mathematics in Foundations. A week before each test, students received an assignment containing about a dozen questions and were allowed to consult their work on these questions during the test. Of the eleven questions on a typical weekly test, four were very close to assignment questions, three or four more were routine modifications of assignment questions, and the rest were unseen questions designed to test deeper understanding and to challenge the more able students. In this way, conscientious students were rewarded for their work on the assignment questions, while high-fliers also had scope to show their paces. Two drawbacks were noted, one educational and one practical:

The multiple-choice tests give students NO PRACTICE at expressing their mathematical thoughts in words and symbols, nor do students receive much individual feedback on their performance in the tests, even though, of course, the correct answers are posted immediately afterwards.

The preparation of the tests (in four scrambled formats to minimise the temptation to copy from neighbours) and the logistics of administering the tests in a lecture-theatre involve the lecturer in a considerable amount of extra work.

I believed that moving the tests onto computers could have the following benefits:

- Quicker, more detailed, and more accurate feedback to both student and lecturer
- Greater flexibility in the question formats and therefore more searching tests
- Increased scope for statistical analysis of test outcomes
- A wider choice of time and venue for taking the tests
- Simplified administration, once test formats and question banks are set up.

With these beliefs in mind, Alyson Stibbard and I applied to the University's Teaching Development Fund for seed money to teach the second-year course Number Theory using directed-learning materials supported by computer-assisted assessment. The bid was successful and it is thanks to the Fund's support that the experiment got under way last year (1997-98).

### **RATIONALE**

A large part of our effort went into the preparation of the directed-learning materials, a set of five workbooks, each between 30 and 40 pages long, designed to help students develop insight into the theoretical knowledge through experiment and calculation, and to discover some of the theorems of Number Theory for themselves. I will not dwell on the details of the workbook method and the philosophy behind it, except to say that students had to do a lot of practical work with their pocket calculators and that the treatment of the material was rigorous and searching. (The previous year I had taught the same syllabus for this 5-week course in 15 conventional lectures.) We adopted the following weekly routine for the course:

- Friday: The week's workbook is distributed
- Tuesday: A one-hour survey lecture to describe the broad ideas
- Thursday: A two-hour drop-in surgery to deal with individual problems
- Friday: (1) Students take test.
- (2) Answers to the workbook problems posted on the Web.
- (3) Next workbook handed out.

I want to stress that the course required students to work very hard in the Summer Term at a time when they were also revising for their core-course examinations. Given that the course was an option and for only 6 CATS, it was surprising that most of the 180 students that started the course stayed with it (172 took the final examination). Given the level of commitment we were demanding, it was essential to provide a framework for regular work and feedback, and this was the role of the weekly multiple-choice tests. In the event, the first two tests were held in the lecture theatre (much like the Foundations tests described above) and the last three were mounted on the University PC network and taken in the Computer suites under the Library on Central Campus.

### **PREPARATIONS FOR THE COMPUTER-BASED TESTS**

Thanks to a grant from the University Teaching Development Fund, we were able to employ an enthusiastic postgraduate student (Benedict Carr) to help with the computer assessment. His brief was:

- to investigate the availability of suitable assessment software and report on its fitness for purpose;
- to adapt the chosen software to the preferred test format and to prepare for implementation on the University Network in collaboration with Computing Services;
- to create and install the files for each test, to monitor the running of the tests, and to analyse the data in the students' answer files after the tests.

These three parts to his brief were carried out in the corresponding terms of the Academic Year. By Christmas we had settled for the software package Question Mark, which had already been tried

out elsewhere in the University. In the New Year we bought two upgrades to the latest version (3.0) and began fruitful discussions with Dr Jay Dempster on educational technology issues and with Keith Halstead and colleagues in Computing Services on logistical questions about hardware, the network, and security. By the end of the Spring Term, a feasible plan had been agreed, and during the Easter vacation a dummy run was prepared. The course began at the start of the Summer Term, but for various practical reasons, we postponed the transfer of testing to computers until Week 23. In the event, three successful tests were taken on PCs by the 175 students then registered for the course.

### **RUNNING THE TESTS FOR REAL**

A frisson of anxious excitement was discernible on the Friday morning of Week 23 when Keith Halstead, Ben Carr, Alyson Stibbard, and I arrived 30 minutes before the start of the first test in order to boot individually each of the 90 PCs on the ground floor of the Library building. We had reserved the only two bookable labs (containing about 85 working machines) and we commandeered a few extra machines from the public area. The students were told to enter their names and student numbers on the first two screens and then to wait until given the word to begin the test. They had received instructions in advance on how to operate the easy test interface, and these were repeated on the back of attendance slips placed by each computer. (The attendance slips proved to be an unnecessary additional safeguard and were abandoned later.) It took at least 5 minutes for around 90 students to file in, find an unoccupied PC, and settle down in front of it. The test lasted 45 minutes and students were expressly asked not to submit their set of answers until the end of the test was announced, whereupon the program immediately returned their scores. (They could navigate around the questions and review their answers at any stage before submission, and the submission button only became active when all questions had received a response.) Hardly any students had problems with the test interface, and in one test students quickly identified a wrongly-posed question where none of the answers offered for selection was valid.

On the first occasion, only 15 minutes had been allowed between the end of the first sitting and the start of the second. In this break between tests, the first set of test files had to be carefully saved, and then all 90 machines had to be rebooted one at a time by hand. In the event, it took four of us over half an hour to prepare the labs for the second wave of candidates, who were naturally impatient with the delay. Because of this, and because of the slight risk of students from the first session passing on details of the questions to their friends waiting for the second, the last two tests in Weeks 24 and 25 were taken in one sitting. Keith Halstead used his authority as Director of Computing Services to close the public areas in order to make all 180 PCs available to run the test in a single 9 o'clock sitting, not a popular time with our number theorists, but one that caused least disruption for the rest of the student body.

### **DIFFICULTIES ENCOUNTERED**

These relate to the strengths and weaknesses of the software Question Mark. First it must be stressed that the program proved very robust in action. In spite of the limitations of the PCs (in particular, their small amount RAM) and the large number of students logged on (over 170 for the last two tests), there were very few problems. Once or twice, in the last few minutes of the test, one or two machines crashed, but the tests files were retrieved (and the students affected were allowed quickly to run through their tests again as an insurance). The crashes may have been caused by students toggling rapidly back and forth through the questions to check their answers before submitting. The students found the program's interface easy to use and needed very little help during the tests.

The program has three components for (i) designing the questions, (ii) running the tests, and (iii) reporting and analysing the results. We found the first of these, Question Designer, the least satisfactory in practice. The test questions were first written in Word and then transferred one by one to Question Designer using the slow and laborious 'copy-and-paste' method. Mathematical symbols and equations could be embedded either with Word's add-on Equation Editor (clumsy and a source of font-size problems) or by snapshotting them as graphics (time-consuming). The ability to input questions in TeX or LaTeX would have greatly reduced the workload of authoring

questions. I also had to produce a differently-formatted version for a blind student requiring special invigilation. Such is the clumsiness of Word with Equation Editor that it took me an extra hour to convert each of these test files. If I had been able to produce the questions in TeX, the reformatting would have taken a matter of seconds.

Further shortcomings include:

(i) The numerical question option allows for only one entry per screen, ruling out a series of linked numerical answers.

(ii) There is no parsing facility to check numerical answers entered in equivalent forms (e.g.  $05 = 50 = 1/2 = 0.499$ ).

(iii) Establishing the user's identity caused problems. The test format offers one initial screen with a box for the student's name but no additional space for the student's University ID number. The screen for the first question had to be used for the ID number which meant that the question numbers at the top of the screen were out of step with the actual numbers of the test questions. It would also have been useful to have had a screen for instructions about the interface and general advice on the test itself.

### STUDENT REACTIONS

Feedback from Course Evaluation Forms and from talking to individual students suggested that the new approach was generally well received. The hands-on method of learning provided by the Workbooks was a new and positive experience for many and the tests were also seen as a key component in providing motivation and feedback. The weekly tests were seen to be fair and at an appropriate level. There were no complaints about the 25% exam credit awarded for the tests (5% for each test), although one or two respondents suggested they would have supported harder tests for more credit. Among the positive comments were the following:

"The layout of the course is ideal, with plenty of examples to do outside and weekly tests to give motivation"

" Workbooks very useful: provide lots of practice at doing examples. Tests also help to firm up ideas."

"Enjoyed new method of teaching. Tests good to help learn stuff."

### EVALUATING THE EXPERIMENT

No formal evaluation has been carried out. The examination results were considerably better than the year before, when it had been clear that a significant number of students had failed to do any serious work on the course. The mean overall score (tests plus final exam) was some 12% up on the previous year's, while the final examination was considered to be harder by the second internal examiner (who had not been involved in the teaching either years). My overall impression of the answers on the scripts this year was of a much better grasp of the material; very few candidates submitted complete nonsense.

The scores of a significant number of students were markedly higher on the tests than on the final examination. I believe this was because a number of students who successfully carried out the step-by-step calculations and deductions in the workbooks were unable to retain a grasp of the broader picture or to reproduce the theoretical material required in the final examination. I intend to write to these students to find out their own explanations for the discrepancies, and to ask how the course might have been modified to help them deal with it better.

### FUTURE PLANS

The framework for testing we used this year incurred considerable costs of time and resources. For this reason and also because of the mathematical limitations of packages such as Question Mark described above, we are keen to explore the possibility of moving to Web-based testing next year. Among the benefits we see are

- Greater flexibility and variety in the types of questions we can ask

- The option of letting students sit the tests over a longer period of time at any machine with a Web browser
- A simplification of the logistics of running the tests
- A reduction in preparation and supervision, once the database of questions is prepared and installed.

We are currently looking at the possibility of using (i) Webtest, a package for mounting mathematical tests on the Web being developed in Heriot-Watt University and (ii) Perception, the Web-based testing package recently launched by Question Mark. Webtest uses XML and has plans to allow questions to be written in Mathematics Mark-up Language (MML) as soon as standard Web browsers can read it. Perception requires an NT server, which is not currently available on the University network. Use of QM Web, the add-on to the version of Question Mark described here, which converts questions/tests to web-based tests, for use on the university's central UNIX web server (running Apache) will also be considered, should a working system be centrally provided.

---

**Dr Trevor Hawkes**  
Mathematics Institute  
University of Warwick

**To discuss any issues raised by this article please contact CAP tel: 024 7652 4766 email: [cap@warwick.ac.uk](mailto:cap@warwick.ac.uk)**