Answer Files --- What more do they reveal?

by

Martin Greenhow (Department of Mathematical Sciences, Brunel University, email: martin.greenhow@brunel.ac.uk)

Introduction

Over the past 5 years, my final year project students and I have written an extensive suite of objective tests called Mathletics. Mathletics now comprises some 4500 questions spanning 175 different skills areas within the GCSE to level 1 university level. The thinking behind the question libraries has been described in Greenhow (2000) whilst Kyle (1999) provides a review of an earlier version of Mathletics. Increasingly, Mathletics has become a workhorse of our teaching at Brunel; last year some 600 foundations, mathematics, engineering and biological sciences students took over 23,000 diagnostic and continual assessment tests, whilst so far this year the usage has been even higher. This paper looks at two student groups and asks what can we learn from this experience, and particularly, what can we learn from the answer files?

Assessment is the driving force for many students entering university today. Put bluntly, many, especially weaker, students will ask "How many marks is this worth?" rather than "What will I learn from this exercise?". Even if they are still attending lectures and seminars, problem sheets are left undone by these students. In contrast, setting regular objective tests with marks counting towards the module total, does at least mean that some mathematics is done (perhaps even learned?) by weak students who see a good set of marks for the continual assessment part of the mathematics module as a hedge against anticipated weak exam performance. At the other end of the scale, good students also seem to appreciate the revision and "fitness-training" that regular testing, with feedback, provides. In their own words: "Mathletics keeps us at it".

Why use diagnostic tests?

Most, perhaps nearly all, university departments requiring a numerate student intake, will be feeling uneasy about the mathematical preparedness of their student intake, especially in the next academic year when so many (29%) students failed AS level mathematics. Whilst this may not affect our mathematics undergraduate intake (predominantly with a good mathematics A level or equivalent) it will certainly affect our service teaching, for example to Engineering and Biological Sciences. Even with a good A-level, there is substantial doubt about the ability of students to apply their knowledge in anything but the most structured of problems. What is needed in both cases, is to ascertain and, if possible, correct, and deficient areas of mathematics right at the start of the first year. For Brunel’s Foundation Programmes, this is even more important given the diversity of the intake.

Reasons for using online diagnostic tests therefore include:

- Productive use of the window of opportunity at the start of the course by keen students with little to do yet! This may, indeed, be the only time when some students will do anything without marks.
- Revision of GCSE and A-level material and retry of the tests. Even if the answer files were not looked at, this aspect would mean that diagnostic tests would still be worth doing.
• Individual diagnostics on a student-by-student basis. Individual profiles are emailed to all students and tutors. This results in a clear action plan for many students who “hit the ground running” when their courses start.

• Formation of a whole class view, whereby bridging sessions can be planned (see below). Cox (2000) claims that by knowing A-level scores, one might infer what skills students know; for example any A-level would imply that the student can use the general quadratic equation, but only those above a certain grade would be able to complete the square. This interesting idea provides a motivation for the data analysis below. However, given the factors above, and the speed with which diagnostic results are returned to students from an online system, Cox’s approach has not be used at Brunel University.

• Formation of a whole class view that informs the level 1 teaching and, eventually the syllabuses, curricula and admissions policy. In particular, in the future it may be necessary to look at intakes with AS level mathematics rather than the full A-level. It will be crucial to ascertain what these students can really do when they arrive at university.

• Linked to all of the above, it is interesting to look at the year-on-year comparability of freshers’ diagnostic tests, see e.g. Hunt and Lawson (1996) and Lawson (2000), and also their continual assessment tests.

Many of the same comments apply to the continual assessment tests. Staff appreciate the automatic marking and feedback, since it leaves them free to do what they do well (talking and listening to students). Online tests can also make the task of providing performance indicators very much easier. For example, the foundation students’ continual assessment profile results show a modest improvement from last year.

**Diagnostic tests results – Foundations students**

The Foundations students comprise typically about 100 Foundations of Engineering (FoE) and about 70 Foundations of Science (FoS) each year. FoE mainly leads on to degrees in traditional engineering subjects, especially Mechanical Engineering. The majority of these students have some mathematical background beyond GCSE and, furthermore, they know that mathematics will become increasingly important for them. This makes them on average stronger in mathematics than FoS students, many of whom do not have any mathematics beyond GCSE and intend to progress on to information technology or biological sciences where the importance of mathematics is less apparent to them. Diagnostics is therefore especially important for FoS.

A 4-year total of about 1000 answer files (not students – see below) from the diagnostic tests are split into 3 groups (poor, average and strong): this provides a mechanism for seeing the absolute and relative difficulty of each tested skill according to the group. This follows the idea of Cox (2000), but from actual data rather than inference from previous qualification. Importantly it informs the sequence of any remedial programme (bridging classes or reference to books or other CAL packages). For example, one could set a threshold criteria/teaching strategy for each cohort; 30% for poor students who should attend bridging classes, 50% for average students who should learn with the help of regular lectures and/or previous mathematics notes (revise their A-levels) and 80% for good students who should learn whatever skills they have not fully mastered independently, perhaps reporting back to their personal tutors. This then adds value to all students, promotes confidence for the average students and underpins mastery for the good students.

As a rough guide, any remedial programme should address the deficient skills from the bottom of the figure up; at Brunel University, many of the lower skills are actually taught as the numeracy part of a study skills module, rather than as part of the mathematics module itself.
A strategy used in 1999/00 and 2000/01 was to encourage retake of the diagnostic tests after revision. This then means that some students will be able to promote themselves to a higher group, thus provide real incentive to address very basic mathematical difficulties they have.

The results from year to year are generally quite consistent, especially for the strong group. To understand the results at the deeper level of tested skill rather than tested topic, see Greenhow (2000), it will be necessary to look at the questions in more detail. For example, why is it that even good students found difficulty with proportionality? A typical question was "y is proportional to \( x^2 \). Given \( y=200 \) when \( x = 5 \), what is \( x \) when \( y=800? \)" Whilst the arithmetic was very simple, the questions were stated in words, writing down the equation was required and, importantly, a two-step process was needed, first to find the constant of proportionality, second to apply this. It is planned to write this type of meta-data to the answer files in future objective tests. The students' own words were "We don't know the formula for this" and so they stopped (most of these questions required a number to be typed in, so guessing was not feasible).
Zero level diagnostic tests 1998/02

- simultaneous eqns
- substitution
- solving simple eqns
- simplify expressions
- indices
- rearranging eqns
- flow diagrams
- factorisation
- expansion of brackets
- coordinate geometry
- expressing simple functions
- Equation Of a Line
- Proportionality
- Generating Sequences
- Percentages
- Powers of numbers
- Fraction, Fractional expressions
- Decimal Scientific Notation
- Using & understanding special term
- General Arithmetic

Group 10-40% average (%)
Zero level diagnostic tests 1998/02

- simultaneous eqns
- substitution
- solving simple eqns
- simplify expressions
- indices
- rearranging eqns
- flow diagrams
- factorisation
- expansion of brackets
- coordinate geometry
- expressing simple functions
- Equation Of a Line
- Proportionality
- Generating Sequences
- Percentages
- Powers of numbers
- Fraction, Fractional expressions
- Decimal, Scientific Notation
- Using & understanding special term
- General Arithmetic

Group 41-70% average (%)

(Marks range from 0 to 100.)
Diagnostic tests results – mathematics undergraduates

The much more homogeneous group of mathematics freshers comprise a 60 strong A-level cohort with grade B or C in Mathematics and about 10 ex-FoS students. They were required to do two diagnostic test in algebra and two in calculus. The total of about 500 answer files are split in the same way as for the Foundations students with the same remedial strategies in mind. The tests have changed slightly over the 4 years, notably that in 2000/01 "series" was
split into "APs" and "GPs", "polynomial multiplication" was dropped because it did not cause significant difficulties, and the two sections on "integrating factors" were dropped because they were too hard. Also, significant was the fact that in 2001/02, the second (harder) tests in algebra and calculus were not compulsory, with the result that only about 5 students actually did them!

Again question-by-question analysis is required to understand the reasons behind students’ facility with each of these skills. However, difficulties seemed to arise from:

- not having studied skills such as determinants at A-level at all (this is useful diagnostic information of course). It was explained to students that they were not expected to be able to do all the questions, and indeed some would only have been accessible to those with 2 mathematics A-levels.
- knowing how to do the question, but not being able to cope with the complexity of what was required. For example, they could expand $(2+x)^5$ but not $(3-t/2)^5$.

A downside to diagnostic testing is now immediately apparent: one sometimes finds out rather too much! Some students in the weak group fail to achieve even the 30% threshold in too many areas. Such students face an enormous, and possibly de-motivating, struggle in the first semester while they must make good the most essential deficient areas (as a rough guide one should again address these from the bottom up; many of the skills at the top will be formally taught later anyway). Is this asking too much of them? If so, are their skills accurately reflected by their entrance qualifications and/or should our admissions policy be changed with respect to these students? In contrast, the average and good students often relish the challenge of brushing up old skills and even learning new ones ahead of their formal teaching. They will often then take a Mathletics test on just these skills (rather than take the whole tests again).
Continual assessment results - Foundations students

The continual assessment tests form an integral part of each semester’s mathematics module, counting 20% for semester 1 and 30% for semester 2. Students are given a deadline at the end of the semester by which all tests must be completed and advised which tests to do on their problem sheets. An interesting feature is that a complete set of marks so far for all students is sent to them by a weekly email (this is done for purely practical reasons, but does add to the peer pressure to perform; your friends will also know how well you are doing!).

Apart from the marks, students know that the same skills will be tested in the exam so they really try to achieve 100%, generally take each test between 2 and 5 times (getting a random selection of questions each time of course). In fact, it is sometimes necessary to ask a student with 95% to stop; getting 100% will not teach them anything more and they will not get the "Grand Wizard" status awarded by PC games. We call it the Nintendo effect.
Zero level continual assessment (00/01 sem 1)
Fig 3 - Average, standard deviation and % usage (from the bottom) for continual assessment tests taken by Foundations students in 2000/01.

Continual assessment results - mathematics undergraduates
The continual assessment tests taken by first year mathematics undergraduates forms 10% of each of the 3 core modules (algebra 1, calculus 1 & 2) in semester 1. Again repeat tests are
allowed, so students seem rather satisfied with their performance. However, even from the crude averages presented here, it is clear that many students will not have achieved a decent level of proficiency in several areas of integration and the use of vectors for solving problems involving geometrical thinking.

Fig 4 - Continual assessment tests taken by mathematics students in 2001/02.

**What difference does Mathletics make?**

The principal answer to this might seem rather strange since it is essentially, nothing to do with the subject matter of mathematics. First and foremost, I see the Mathletics activity described above as a mechanism for setting up a "social context" in which learning can take place. The relaxed atmosphere of the PC lab encourages group work, student-to-student teaching, peer pressure to get full marks for each test and a strong sense of progress with, and reward for, clearly defined and accessible, but not trivial, tasks. The PC labs also provide many ideal
opportunities for one-to-one teaching since it’s "you and the student against it" rather than the more usual role for the teacher as task setter and marker i.e. "you against them". Finally, students appreciate that your eye will be on their answer files and advice issued (via email) from time to time. This may lie behind a comment I had not expected: "Mathletics provides a link between the students and the lecturer".

Common sense suggest that learning must be taking place; sceptics would point to the lack of a control group, small sample size, the confounding factor that only interested (interesting?) lecturers get involved with CAA in the first place, the self-referential nature of any study showing improvements (perhaps they are just getting better at doing objective tests, not mathematics), and much else besides. However, an accepted comparison might be with the unseen written exam at the end of semester 1, see Fig 5. Given that it is always possible for students to improve their overall Mathletics mark, it is not surprising that Mathletics over-estimates their true mathematical skills in general. Note that there is considerable scatter (especially for the less homogeneous FoS cohort); some students with a strong mathematics background will not bother to improve their Mathletics score or even complete all the assigned tests since they know they will pass the exam anyway.
Fig 5 - Exam/Mathletics comparison for FoS and FoE 2000/2001 students. (Note that 0% for the exam indicates absence.)

Conclusions
The above programme of continual assessments via online objective testing is felt to be a highly motivational educational strategy that complements traditional lecture courses and paper-based exercises and exams. However, uninvigilated and repeatable-on-demand online objective tests do not appear to rank students correctly. Traditional exams are still needed, not least to provide impetus to students to complete the required menu of continual assessment tests.

References