[O8] Learning via online mechanics tests: update and extension

Mundeep Gill and **Martin Greenhow**
Department of Mathematical Sciences, Brunel University
acsrmkg@brunel.ac.uk; mastmmg@brunel.ac.uk

Keywords: objective questions, mechanics, error taxonomy, evaluation

Abstract

This paper describes the use and extension of online mechanics tests written as part of the PPLATO (2005) project and described at the last S.L.T.C. by Gill and Greenhow (2005). By using random parameters within questions, millions of question realisations will be seen by students, each with fully-worked feedback, including run-time-generated diagrams if required. An implication of the randomisation is that students' answers (numbers or choices) recorded in the answer files need to have associated metadata that tags their meaning. This meaning requires an overarching *taxonomy of errors* to identify the class of error being made by the student (procedural, conceptual etc). We will describe the specification of this taxonomy and the relative frequency of errors being made by the last five years mechanics classes. We show how this data can be used to inform edits and updates of the questions by enhancing the (hidden or displayed) distracters using evidence-based *mal-rules* as well as those arising from *break points* in the question's solution. In turn, this informs the contents of the feedback so that we are able to suggest *where* the student went wrong, rather than just that they were wrong.

We report on a continuation of our FAST (2005) experiment designed to study the efficacy of the feedback. The use of *indicators* in subsequent exam scripts supports the view that students' retention period of the techniques given in the feedback and their ability to answer unseen exam questions correctly is linked to their engagement with and performance on (generally repeated) tests.

Background

The Mathletics system currently comprises some 1600 *question styles* (see below) that exploit the use of random parameters within highly-structured questions that test the mathematical and statistical skills of GCSE students to second level undergraduates. In particular A-level content in algebra and calculus (C1-C4 modules), statistics (S1) and mechanics (M1 and some of M2) is now quite mature and in heavy use at Brunel University with our Foundations of IT and first-year mathematics students. Although based on A-level syllabuses, the material overlaps significantly with that taught at university in a range of disciplines that require mechanics (e.g. engineering, physics, sports science). That some Brunel students find this material difficult is almost certainly due to the changes in the A-level curricula, whereby they have had less exposure to the application of mathematics, rarely going beyond the M1 syllabus, and often not even that, preferring to take their two applications modules form statistics and/or decision mathematics. The inexorable slide in numbers taking A-level physics has further compounded this situation. We believe this situation is not unique to Brunel University, but commonplace across most of the HE sector. We are left with a significant need to diagnose and cover elementary material again during the first half of a typical level one undergraduate mechanics module, before progressing to more advanced material. We present evidence that the Mathletics question database, used for formative assessment with rich feedback, is very helpful in this task. Our experience over the last 5 years of trials with many hundreds of students indicates that they value the extensive feedback as a learning resource, as well as for the marks awarded. Thus student engagement and time-on-task is generally assured, which builds confidence, having a beneficial effect on other learning tasks within the module. The pedagogy of building tests into a module is quite well established, and various trials of the mechanics material (described and updated below) have indicated that students move, at least partially, to a deeper approach to study (Gill and Greenhow, 2005).

The underpinning technology, whereby many thousands or millions of question *realisations* are generated by a single question *style* that encodes the algebraic and pedagogic structure of the question, is extremely helpful in moving students away from simple memorisation towards understanding the question's content and solution. The random parameters, possibly constrained according to the question's content (realism of the question and reverse engineering from a desirable solution form), are carried through to all parts of the question so that it realises with:

- dynamic MathML, giving equations in the question and in the (often extensive) solution and other content given as feedback.

- dynamic SVG, giving accurate diagrams, charts and graphs.

- dynamic wording, giving different scenarios, expressed in gender- and ethnically-balanced language.

- dynamic question functionality, such as algorithms that, when run to completion, generate, for example, HTML tables of variable length.

Other technical issues, described in Ellis *et al* (2005), include:

- Accessibility. This has been a key feature of all elements of the questions. Student preferences are stored and used to resize/recolour all text, equations, diagrams and tables.

- Functions. A great deal of technical effort has also gone into the writing of functions to underpin the questions. These split into two basic types: functions that return the result of a calculation and functions that return display strings e.g. a MathML string to display a curl in determinant form or an SVG string to display a force diagram or a speed/time graph.

- Exportability. All of the above generic issues concerning the display of mathematical content on a web page may be exported to other web-based CAA or CAL systems.

A taxonomy of errors

Whilst the overwhelming utility of using random parameters within questions is clear, there are a number of consequences for question design. We elaborate here on how (displayed or hidden) distracters within questions can be attributed sensible metadata to be recorded in the answer files. Our approach to the design of distracters is based on the belief that

Error Type	Classification
Generic errors that may occur in most/all areas of mathematics	
Assumption	Students assume certain things that are not true, for example, in projectile questions, that vertical velocity is equal to initial velocity.
Calculation	Method correct but calculation errors are made.
Copying Errors	Copying values incorrectly.
Definition	Not knowing the definition of terms given in question text, e.g. magnitude.
Formulas	Incorrectly stating/recalling formulas.
Incorrect Values Used	Using incorrect values in method, for example, when substituting values into formulas.
Knowledge	Knowledge students are lacking that would enable them to answer questions.
Methodology	Students attempt to use an incorrect method to answer a question.
Modelling	Generic definition, e.g. ignoring forces acting on particles for example, gravity.
Procedural	Method student attempts to use is correct but can only do initial/certain stages of the method. They stop halfway through when they do not know the stages that follow or when they are unable to interpret initial results.
Reading	Reading the question text incorrectly and confusing the value of variables.
Trigonometry errors	Basic definitions of cosine, sine and tan incorrect. This is most apparent in questions where students are required to resolve forces.
Resolving Forces	Resolving forces incorrectly (specific to mechanics)

Table 1: List of errors and their categorisation

most wrong answers (from students who are trying rather than guessing just to study the resulting feedback) in either CAA or paper-based tests are not arbitrary, but result from logically-structured but incorrect methods of solution. We have therefore spent much time identifying such *mal-rules* from past examination scripts and encoding them in questions. Mathletics then is able to recognise selected answers/input numbers thereby offering students targeted feedback that tells the student not just that they are wrong, but *why*.

This is, of course, very beneficial for the student; for staff, the answer files need to record metadata for student choices (outcomes) that reflect more generally and usefully how students are making errors. For example, the three-parameter question style with descriptor ‘ax+b when x = c’ might realise as ‘What is the value of -5x-6 when x = -2?’. Recording an answer of -16 is less useful that recording how that answer might arise; the outcome metadata must therefore somehow encode the information that the student

Error Type \ Topic	Weighted Mean: Vectors	Kinematics	Dynamics	Statics	Overall Weighted Mean
Calculation	2.86%			3.62%	**2.203%**
Copying	0.58%	1.27%		0.27%	**0.365%**
Formula	5.18%			0.73%	**1.861%**
Incorrect Value	0.29%			1.67%	**0.699%**
Knowledge	6.32%	16.50%	2.03%	14.11%	**8.764%**
Methodology	1.13%		6.53%	2.20%	**2.836%**
Modelling			0.34%	0.73%	**0.353%**
Procedural	1.46%	12.65%	6.22%	7.24%	**5.578%**
Reading				0.23%	**0.084%**
Trigonometry			2.75%		**0.709%**
Correct	70.73%	63.18%	42.60%	49.81%	**55.210%**
Unable to Answer	11.44%	6.38%	39.50%	19.38%	**21.210%**

Table 2: Table showing the weighted mean for each error type that occurred across the four subtopics and the overall weighted mean for errors that occurred at level one

cannot handle double negatives. The same outcome metadata might pertain to other questions too, meaning that the student genuinely doesn't understand, rather than that (s)he simply made a slip (also serious!). For simple algebra, strictly formulaic mal-rules appear to be useful; for the mechanics questions described here, value judgements need to be made, especially on what can be assumed about the student's elementary algebra and calculus skills that are needed as small parts of a larger calculation. This means that a more abstract categorisation of error type is needed in order to understand individual students' or whole-cohort profiles, as given in **table 1**. These error types arose naturally from the analysis of five years' worth of examination scripts (126 in total) and are used to:

- summarise the relative frequency of errors in the four main mechanics areas for which CAA questions were set and run, see **table 2**.

- design those questions, especially the mal-rules and their associated metadata, see **figure 1**.

Efficacy of feedback

Anecdotal and questionnaire evidence strongly indicates that students engaged with, and even enjoyed, a programme of structured and staffed lab sessions during the 2004/05 and 2005/06 academic years. Although these sessions did not count towards the overall module mark, they were explicitly built-in to the module timetable and highlighted by the lecturer. It is natural to ask how one might measure the learning effects of the feedback, for how long do such effects last and are they applied out of context, even in other unrelated modules? To attempt to answer the first two questions, we looked at (rather mundane but easily identifiable) *indicators* in their lab workings and, as shown in **figure 2**, their end-of-module exam scripts. The indicators reflect features of the feedback and we seek to see if students mimic them by: using properly-constructed diagrams (early) in their solutions; laying out their methodology in a clear way; quoting units in their final answers; underlining vectors. **Figure 2** shows a clear increase in the use of diagrams in the period

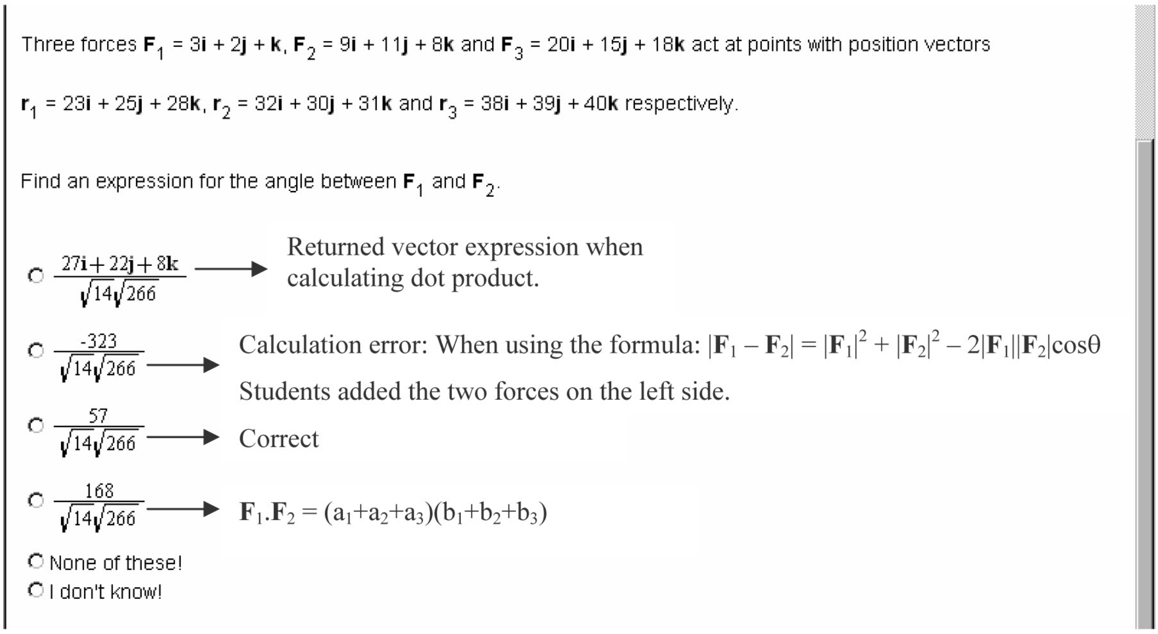

Figure 1: A typical multi-choice question. The encoded mal-rules (in blue) arise from the analysis of break points in students' working (i.e. points where students' working commonly goes wrong) and their metadata is recorded according to the error categories of **table 1**

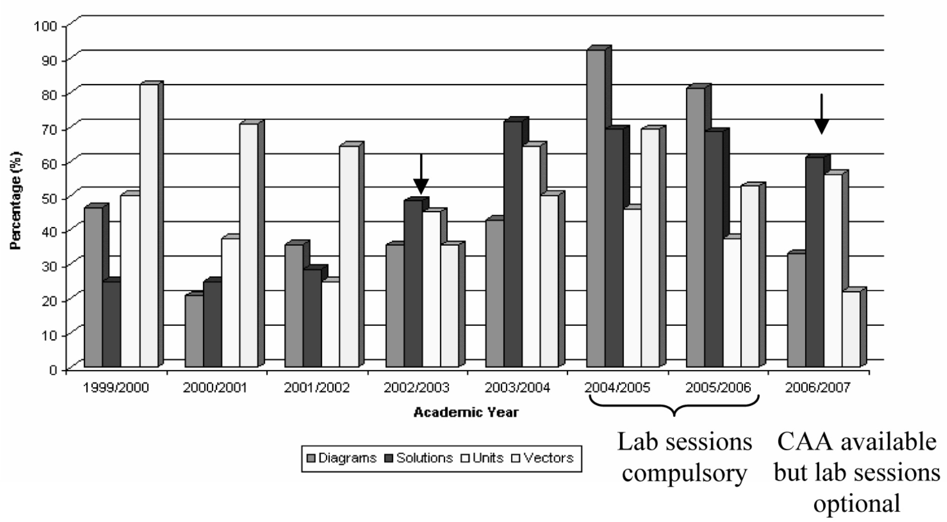

Figure 2: Indicators showing the effect of CAA on students' behaviour in subsequent exams. Vertical arrows indicate a new lecturer taught the module

of 'compulsory' CAA, some improvement in solution layout and identification of vectors, but students still seem impervious to efforts to get them to use units. What is new to this figure compared with Gill and Greenhow (2005), is this years' worth of data (exams took place in January 2007). **Figure 2** reflects the fact that very few students tried the (now optional) CAA tests, none significantly, which means that CAA of this type needs to be built-in (possibly for marks) rather than added-on to an already busy timetable. The results, however, do act as a benchmark of the existing situation, see the remarks in the introduction, and suggest that the use of compulsory use of CAA for two years buoyed up the slowly-sinking performance of the student cohort.

Conclusions

The design of objective questions for mechanics problems that typically require 5-10 lines of equations, is facilitated by understanding content-specific mal-rules by which students make errors at break points within their solutions. It is possible to categorise these mal-rules by an overarching taxonomy of errors that may pervade mathematics and other related disciplines. We can then inform and understand what information is recorded in the answer files of any CAA system, especially if the questions fully utilise the potential of random parameters. The method has recently been extended to more advanced topics, such as vector calculus and examples will be presented at this conference.

Much of the effort involved in authoring is spent on writing extensive feedback screens. Students use these as a primary learning resource which they perceive to be of value. We extend the indicator-based evidence presented by Gill and Greenhow (2005) which suggests that students' competence with the mechanics subject material is enhanced, providing that they actually engage with the questions. In practice this may mean that a small percentage of marks needs to be awarded for the CAA tests.

References

Ellis, E. *et al* (2005): Exportable technologies: MathML and SVG objects for CAA and web content. Proc. 9th CAA Conf, Loughborough, July.

FAST (2005): Formative assessment in science teaching
http://www.open.ac.uk/science/fdtl/

Gill, M. and Greenhow, M. (2005): Learning via online mechanics tests. Proc. Science Learning and Teaching Conf. Warwick, 27/28 June.

PPLATO (2005): Promoting physics learning and teaching opportunities
http://www.rdg.ac.uk/AcaDepts/sp/PPLATO/publish/