RESPONDING TO STUDENT EXPECTATIONS FOR ASSESSMENTS

Richard A Bacon

Responding to student expectations for assessments

R A Bacon,
Department of Physics, University of Surrey,
Guildford,
Surrey. GU2 7XH
r.bacon@surrey.ac.uk

Abstract

In a previous paper (Bacon 2003) the author described some of the results obtained from a survey of the use of the SToMP testing system for coursework assessment of a first year Data Handling course within a Physics Degree programme. This paper will deal with the modifications that have been made as a result of the student feedback from that trial, with a preliminary analysis of the feedback obtained from the students using the updated tests, and with a further trial of a more sophisticated response to the feedback.

The SToMP testing system was written to be direct implementation of the IMS-QTI v1.2 specification, but includes several extensions for handling numeric questions of the type most frequently found within science and engineering. Such questions typically require a numeric answer to be judged by its precision (e.g. the number of significant figures) as well as its accuracy (i.e. whether the value falls within a specified range) and to be able to recognised alternative forms of the same value and precision in scientific format. These features were mapped onto an extension of the QTI specification for ease of implementation, and included other features such as alternative number bases and the randomisation of values within questions.

The questionnaire that was used in the earlier paper has been used again with the current cohort of students using an updated version of the testing system and with some questions having been modified in the light of earlier feedback. A major objection voiced by the students in the first trial was the lack of ability to show working, and to get credit for correct working if the final numeric answer was incorrect. This has not been directly addressed within these tests, but an associated problem has been addressed. This concerns multi-part questions and the awarding of marks for later parts that are not correct, but are consistent with earlier errors. The paper will discuss some of the issues relevant to this feature, and how it was perceived by students.

Results will also be presented from a second study into how correct working for incorrect answers to numeric questions can be credited, to some extent. This system allows numeric expressions to be entered by the students as well as their final numeric value. It has been implemented in a test for second year

students on a course concerned with radiation detectors. Clearly this is not being represented as a total solution to the problem but it can be seen as a first step towards an acceptable solution. The student responses to this development could be of particular significance, since they are the same students whose responses have been reported in the earlier paper and that led to the changes being reported here.

Introduction

The SToMP QTI assessment system was used with students for the first time in the autumn of 2002, as described in an earlier paper (Bacon 2003). The system is IMS-QTI V1.2.1 compliant (IMS 2003), with extensions to support question cloning and the scientific use of numbers (Bacon and Smith 2003).

The feedback obtained from students during the first year of use suggested that questions requiring a single numeric value as an answer (i.e. a typical science problem) were seen as being marked unfairly. There are several issues that affect this perception, and some of them were addressed during further development work in the summer of 2003, before the system was used again. In its second year it was again used for the first year data-handling course (nine tests with from two to eight questions each) and a second year seven question test about radiation detectors. It was also used with two first year electronics tests, although this use is not reported here.

The overall aim of this work is to transfer components of assessed coursework from paper to computer in a manner that is acceptable to students used to submitting such work on paper. It is recognised that there will always be a minority group who do not like using computers at all, and another minority group who particularly like using computers. This work is really aimed at the third group (which is by far the largest) comprising those who don’t mind how they do such work, so long as they receive fair reward for the effort they put in, in a transparent way.

New features

Error propagation

Two new features have been implemented in the SToMP testing system and used with students during the 2003/4 session. The implementation of the first of these was described in detail in the earlier paper (Bacon 2003[1]), and so will not be repeated here.

This first feature is applicable to numeric questions having several parts, and allows the system to generate alternative answers to subsequent parts of a question if a student enters an incorrect value for an earlier part. The incorrect value is used as the basis for the alternative answers to later parts. The purpose of this feature is to stop an error on the part of the student propagating through a sequence of related calculations.

Consider, for example, a question asking the student to find an average of twenty numbers. This could be split into two parts with the first part asking for the sum and the second part asking for the average itself. If the student calculates the sum incorrectly and then enters an average based upon that incorrect sum, this feature will allow the value to be recognised as the student's value for the sum divided by the number of items. Since the student has demonstrated knowledge of the method, full marks might be given for the second part although no marks will be given for the first part.

A concern here might be that if the student knows that this facility is being offered, they might deliberately enter a wrong value for the first answer that offers a simple calculation for the subsequent answers. It could be argued, however, that this is in fact giving credit correctly. The student will not get marks for the part of the calculation they did not know how (or could not be bothered) to do, and they do get marks for showing knowledge of the subsequent method. It is up to the author to make sure that the marks awarded to each section accurately reflect the their worth.

This feature was implemented in one set of eight questions and another set of seven questions in the first year Data Handling tests and one set of four questions in the second year Radiation Detectors test.

Numeric expressions

The second feature supported the entry and marking of numeric expressions. This was a first step towards assessing students' working in numerical problems.

In order to allow the entry of an expression, a new response type was defined called <response_exp>. This element is appropriate for use with the <render_fib> interaction element for numeric expressions and with a prototype element called <render_exp> which displays an algebraic expressions editor. This latter prototype element was not used in the work described here - only numeric expressions could be entered.

The entering of numeric expressions was introduced in the general instructions for the test:

> Numeric expressions should be entered in a programming style (e.g. 3.7e-6 * (1- (0.99^2) / ln(27))). Negative values should be enclosed by braces.

Each question offering this feature also displayed the rubric:

> The expression may contain the operators **+ - * / ^**, matched braces **( )**, the functions **ln( )** (natural log), **log( )** (base 10 log), **sqrt( )** (square root), **exp( )** (power of e), **sin( )**, **cos( )**, **tan( )** and numeric values.

Expressions were evaluated by means of a SToMP project expression evaluation routine and the values were checked for accuracy against the same question variable (CETIS 2003) that was used for the answer value. The expression was checked using the <questvar_equal> element that formed part of the SToMP numeric question support (CETIS 2003), but with a new "expression" attribute. An exact match would be most unlikely and so a match within +/- 1% was considered correct, and one within 10 or 20% (depending upon the question) was considered inaccurate and scored part marks.

The marking scheme used with these questions was introduced in the general instructions:

> Some numeric questions allow you to enter a numeric expression as well as your final value. In these cases a correct expression will only contribute to your mark if your final value is wrong.

and in each appropriate question was described as:

> Half marks will be awarded for a correct numerical expression.
> Full marks can be obtained by entering just the correct final value.

Student feedback

The first year students were given exactly the same questionnaire as the previous year, and 26 responses were obtained from the 56 who took the tests. The second year students were sent an abbreviated email questionnaire, and this elicited a similar proportion of responses - 19 from the 38 students who took the test. The following tables summarise opinions expressed by four or more students for the second part of the questionnaire. This requested free text responses to five questions about the best feature/worse feature, reasons for/against using such e-assessment, etc.. The topics of the responses were not suggested in the questionnaire.

Responses from first year students

	Number of students	point being made
a	18	Liked the flexibility in timing of the taking of the test
b	16	Did not like being unable to show working and getting marks for it if their answer was wrong
c	11	Liked the lack of time limit
d	9	Experienced system problems
e	5	Did not like having to take each test in one go
f	5	Found the system easy to use

Responses from second year students

	Number of students	point being made
g	16	Did not like being unable to show working and getting marks for it if their answer was wrong
h	6	Liked the lack of time limit
i	6	Liked being able to stop and restart the test
j	5	Liked the open book nature of the test
k	4	Liked the way it encouraged some useful revision

Comment *d* was partly due to problems associated with a move from a dedicated departmental PC lab to a shared (school) laboratory of hybrid Linux/NT machines and partly due to poor network access error handling and reporting by the SToMP system. This latter problem has now been addressed.

Some of the other comments seem to be based upon faulty premisses. The lack of time limit (*b*) was not strictly true - there was a deadline by which the test had to be completed, and so the amount of time spent on the test was only as uncontrolled as if it had been set on paper.

Similarly, the "open book" comment (*j*) seemed to ignore the fact that paper based coursework would also have been the same in this respect.

It was encouraging to find that some recognised it as useful revision (*k*), but again it was no more so than paper based coursework would have been.

Comments *e* and *i* can be reconciled, because the two tests were not administered in the same manner. Following the same comment as *e* in the previous year's feedback, the system was modified to allow the test to be interrupted and resumed at the discretion of the course lecturer. Thus, the first year tests could not be interrupted, but the second year test could be interrupted.

Discussion

The most frequently expressed complaint is that students were not able to submit and gain marks for working (*b* and *g*). There is a strong tradition (amongst undergraduates at least) that the final value of a problem is far less important than the way in which it was obtained. There is a some truth in this and when such work is being handed in on paper a frequent lament by staff (particularly at school level) is that marks cannot be awarded for a problem with a wrong answer if no working is shown. Having trained students to expect at least some marks even if the answer value is not correct, it is perhaps unreasonable to deny them these marks just because the problems are being submitted and marked electronically.

The feature that allowed for the entry of a numeric expression that would be marked if their final answer was wrong, was introduced as a first step towards assessing the working in a problem. An expression encapsulates such working, but in this first implementation only the calculated value of the expression was assessed, not the form of the expression. In practice this aided only two students, and gained them each half the marks available for the question.

With the one question-set where it was most appropriate, the feature that stopped errors propagating through a multi-part question was successful, and helped 20 out of 56 students. This is a particularly interesting result - the task of re-calculating questions according to students' errors by hand is laborious, so in doing this the marking system saved a considerable amount of work, and it did it objectively and accurately. The second set of questions in the first year test to which the feature was applied were rather harder, and there were therefore fewer examples of error propagation and only six students were helped here. In the second year test, the feature was applied to one set of six questions, but the relationship between the questions was less well defined, and no student benefited.

Most students seemed unaware that any marks between full marks and zero were ever awarded for numeric questions. Students were not told about the error propagation protection partly because of a misplaced fear that it would be abused and partly because the instructions for the tests were already too long. In the second year test only one student commented that the numeric expression gave them the opportunity of showing something of their working.

Observed student experiences

Anecdotal evidence indicates that students spend a long time in front the computers agonising over numeric questions. This is not necessarily a good learning experience. One way of alleviating this might be to hand out the numeric questions on paper beforehand, but with symbols instead of the (randomised) numeric values. The required methods can then be revised and perfected before the electronic version is started.

A problem two or three students encountered was that when they entered an expression incorrectly, the error reporting by SToMP did not give a specific message for their syntax error. Neither did it indicate where in the expression the error had been detected. For example one student entered the expression "dx/dt=(0.9)*((2.3)/(3.5^2))" but because the feedback was inadequate, he was not able to work out that the "dx/dy=" was redundant. He eventually submitted this expression, even though the system told him it was incorrect. This feature is now in the process of being improved, with the position of the error being indicated in the error message.

It is perhaps interesting to note that in discussions with students about the use of the system, many of them implied a complete lack of confidence in their ability to transfer a number accurately from one place (e.g. a calculator screen) to another (e.g. a computer via the keyboard). This helps explain, perhaps, their unease with such a high proportion of a question's marks lying with the accuracy of a single number.

Conclusions

It is apparent that students attitudes towards the testing system are being based upon insufficient information about how it is being used in their courses, what it is replacing, how it operates and how it awards marks. Some of this information is outside the scope of the testing system, and might be provided by the course tutor. The way the tests operate, however, and how marks are awarded are both relevant to the system and this information needs to be available to students at the time they take the tests. Experience has shown that students do not even read the whole of the instructions on the first page of each test, however, and so how this information is to be brought to their attention is not clear at present.

Otherwise, however, the error propagation protection feature was considered to be a success, but the numeric expression entry feature really was just a first step towards being able to assess a student's working in a problem as well as their final answer.

References

Bacon R A "Assessing the use of a new QTI assessment tool within Physics" CAA 2003.

IMS V1.2.1 IMS Question and Test Interoperability. Version 1.2.1 Final Specification <http://www.imsglobal.org/question/index.cfm> (March 2003)

Dick Bacon and Graham Smith "Adaptation of Computer based Assessment schemes to meet student expectations". ALT-C 2003 Conf. Sheffield. September 2003.

CETIS (2003) *Question variables (Dick Bacon, University of Surrey)* <http://ford.ces.strath.ac.uk/QTI/working_papers.html> (May 2003)