A National Database Test Server

Jon Sims Williams[♣]
Mike Barry[♣]
John Maher[♠]

[♣]Department of Engineering Mathematics
[♠]Department of Chemistry
University of Bristol

Abstract

A computer system from which students in every university can take tests over the Internet is described. The system provides a database of questions, at present mainly in mathematics and chemistry, from which lecturer can draw to form their tests. The classification system, that allows lecturers to remotely specify tests, is described for the case of mathematics. It is proposed that, with a very large database of questions, Open-Book Open-Access tests are satisfactory and that this saves large amounts of staff time.

Motivation

Higher education is continually trying to maintain and raise standards. This is particularly true of mathematics in view of the shortage of A level mathematics candidates and the expansion of higher education. If we are to raise standards we must provide support to students and this normally is expensive in staff time. A new approach is required that will provide support to students, will help to raise or maintain standards and yet will be cost effective.

The maintenance of standards of mathematical competence requires that students know what is required. This can be done through the use of end of year exams, but this is too late; we need students to know that they have problems and then to correct these before the end of the year. Increasingly students are being given progress tests to help provide motivation and support. Testing students is expensive in staff time. If it takes ten minutes, per student, of a lecturer's time to set-up, administrate, mark and provide feedback on one test, then three hundred students taking four tests a year cost 200 hours of staff time. This paper discusses work that has been done to remove much of this labour by setting up a nationally accessible Database Test Server, TAL (Test & Learn).

The major components of the work involved in testing students in a fully paper-based system are:

1. generate the questions for the test,
2. organise the questions into one or more tests,
3. produce the paperwork,
4. schedule the tests
5. organise the examination sessions and staff to invigilate,
6. invigilate the test or tests,
7. mark the scripts,
8. Collate the marks into a mark sheet and providing these to the university mark system.

If you use a commercial computer-based testing system, you still have to do 1,2,4,5 and 6.
In this paper we propose that by using the TAL system you can potentially only have to do 2 and 4 both of which can be done, with a few clicks of the mouse, from the comfort of your armchair! We now examine this proposition.

The Questions

We have built a system that allows lecturers to set tests using questions from a database. At present the database contains about 1200 mathematics questions, see Table 1, that have been tested on students and we have data on how long students take to do the questions (Time-to-do) and how hard they find the questions (the Facility). Potentially this saves you setting your own questions but , We presume, self-respecting lecturers will want to set some of their own. When these new questions have been debugged, classified and have some statistics, they can be made available to the rest of

the community. This is the justification for saying you don't *have* to do "1 generate the questions for the test".

Subject	MAIN TOPIC	Number
System	1. introduction	5
Mathematics	2. introductory mathematics test	30
Mathematics	diagnostic mathematics	47
Mathematics	algebra	4
Computer Science	c++	19
Chemistry	classifications omitted to save space	1200
Mathematics	complex algebra	97
Mathematics	differentiation	236
Engineering	classifications omitted to save space	2
Mathematics	functions	56
Mathematics	integration	133
Mathematics	maple	54
Mathematics	matrix algebra	168
Mathematics	numerical methods	1
Mathematics	optimisation	1
Mathematics	ordinary differential equations	36
Mathematics	partial differentiation	28
Road safety	classifications omitted to save space	10
Mathematics	vector algebra	216
Mathematics	vector calculus	150
	TOTAL	2584

Table 1 The numbers of different sorts of questions in the database

Classification of Questions

A large database of questions demands a good classification system to make it possible to find the right questions. When classifying a question the focus is on "What did the person who set the question intend to test?". In future we hope that question setters will specify this at the time when they are setting the questions, but at present this task is mainly done retrospectively. The system of classification is illustrated using the scheme for mathematics questions as it is most advanced. The attributes used are listed in Table 2.

Name of data attribute	Definition or example value
Facility or difficulty	% of cohort that can get the question right.
Time-to-do	average time taken to get question right + one standard deviation.
Subject area	maths, chemistry, C++, astrophysics, road safety.
Main Topic	differentiation, integration, vector calculus.
Sub-topic1	max, min or turning points.
Sub-topic2	simple algebraic functions.
List of other sub-topics	unit-conversion.
Theoretical	$\in \{0,1,2 \ldots 9)\}$ the theoretical sophistication of the question.
Style	the style in which it is set may make it suitable for a particular student group. e.g. Aeronautical Engineers, Pure mathematicians

Table 2 Classification Schema for Questions

Consider now the following question:

> Question 2136:
> A solid right circular cone has height, h = 20 metres and density proportional to distance from the base. Find the distance of its centre of gravity from the base.
> {A right circular cone subtends a right angle at its apex from all diameters of its base.}

This is a mathematics question which tests whether the student can construct a simple mathematical model involving an integral over a volume. The model is a centre of gravity model. Only simple algebraic functions need be integrated and the integral is numerically evaluated. So Q2136 has the following classification:

Subject area	mathematics.
Main Topic	integration.
Sub-topic1	simple model construction
Sub-topic2	volume integral
List of other sub-topics	centre of gravity, cones, simple algebraic functions, evaluation
Theoretical	1
Style	non-specific

Table 3 An Example of the classification of a question

For each Main Topic a list of approved sub-topics is available and this is used to constrain the proliferation of terms. The lecturer when setting his test uses the classifications in two ways:

- Subject Area, Main Topic, Sub-topic1 and optionally Sub-topic 2 are used to specify what sort of question is required;
- Subtopic2 and the list of other subtopics may be used to exclude questions from the specification that the lecturer does not wish to use.

Thus if the lecturer wanted to set a question on volume integrals but had not discussed centres of gravity, he could exclude this and similar questions.

This example brings out the need for sub-topics that specify:

- what sorts of functions are involved
- if special geometric knowledge (e.g. cones) is required

Each Main topic has its list of relevant sub-topics but there are also some general classifications of types of questions. Thus, when setting a test, the lecturer may want a question to motivate the student or to test his ability to generate or use simple models. To provide this flexibility we have defined the following sub-topics:

Simple Model Construction - a Simple Model must be generated to complete the question;

Model Application - a Model is provided as part of the question and must be used . This does not imply that knowledge of the application area is required;

Practical application - a Practical Application of the Main Topic to some application area. The application area for a particular question will always be specified as a sub-topic. A lecturer can thus choose to have a Practical application but exclude certain types of application that are unsuitable e.g. Statistical Process Control. Practical application **does** imply that knowledge of the application area **is** required as **no** model is provided. The practical application should have some motivational benefit.

Simple Model Construction, Model Application and Practical Application would usually be used deliberately as part of the design of a test i.e. “We want to include a practical application question in our test”. Thus when a question is classified these will be inserted as Sub-Topic1. The difference between simple model construction and practical application is that the model, in the model construction, is assumed to part of general knowledge rather than a special application area. For example a question that requires one to find the number of times that a cylindrical roller rotates as it rolls the length of a path, requires a simple model to be constructed, but is hardly an application of a set of special knowledge about rollers. Question 2136 is classified as a Simple Model Construction, rather than a Practical Application, because the question is hardly practical and so does not fulfil the motivational role. This rather subtle distinction is typical of the classification problems that exist.

The values of time-to-do and facility for each question will depend on the particular group of students and what they have been taught. So each lecturer registers his own students and has a database table with the values of time-to-do and facility for each question and for each course unit that he teaches or examines. This is essential since a straightforward question for a student, who has seen something similar, is very much more difficult, and therefore time-consuming, for a student who has never seen anything like it before.

Organise the questions into one or more tests

To set a test the lecturer goes to the TAL website (www.tal.bris.ac.uk) and logs on. His username identifies that he is a lecturer and gives him access to his own private question space and the database of public questions. Using the Test Compiler the test is specified in much the same way that one plans an exam. The questions in the database have an average length of about 4 minutes so a thirty minute test will usually need eight questions. In Table 4 a skeleton design for a four question test is displayed.

Example test Design

Test Name: test1 Course Unit: EMAT1011

Setter: Jon Sims Williams Dates available: 3/5/99-15/5/99

Question No.	*Main Topic*	*SubTopic1*	*Subtopic2*
1	functions	drawing graph of function	
2	functions	domain and or range	
3	differentiation	practical application	simple algebraic functions
4	differentiation	function of function rule	trigonometric functions
length of test:	11:34minutes	+/- 2 minutes	
Facility of test	60 %	+/- 5%	

Table 4 - A simple specification of a mathematics test.

This test specification is for a four question test - tests as short as this are unlikely to provide reliable results - however it illustrates the ideas. Each question is defined in terms of three levels of data: {Main-topic, sub-topic1, sub-topic2} which is represented by a Topic-ID. When the user specifies a question like this the system will go and look up all the questions in the database corresponding to this Topic-ID and return a list of additional topics or sub-topics that this complete set of questions would require. For example Q779 is a question on Functions in which the student is asked to draw a hyperbolic function. This question satisfies the specification for question 1, but its sub-topic2 is hyperbolic functions. The system adds hyperbolic functions to its list of other-sub-topics for question 1; the user can then define which of these should be excluded. The system provides the user with the “length of the test”. This is the sum of the average length of questions for the specified Topic-IDs. When tests are generated they are only accepted if the sum of the time-to-dos lies in the range 11.34 $\pm$2 minutes. Similarly all the tests must have an average facility of 60 $\pm$5%. These values can be changed by the lecturer so that if he wants an easier test he can specify a facility of 80% or alternatively he may not mind what the length of the test is, provided it is less than half an hour and more than 10 minutes. He would then specify the length of the test as: 20 $\pm$10 minutes. Since this length of test is considerably greater than that given as the mean for the set of questions specified, the number of tests generated might be very small and he would be better advised to add more questions to the specification.

When choosing the sub-topics for each question it is important to make sure that the number of questions available on the sub-topic is larger than the number of tests you wish to set. If there are not enough questions on the chosen sub-topic then alternatives can be offered. At Bristol we always aim to have more than 20 equivalent but totally different tests. In this way the probability of two students, sitting next to each other, having the same test is low. The order of the questions is randomised at run-time so that even if this

should happen students are not likely to notice. As a result we are able to allow students to take the test in a public terminal room but without direct supervision. This style of testing may not be appropriate for all situations; users can specify that they only want to set one test and have the freedom to inspect the questions in their tests and swap questions in and out at will.

Methodology for Administering Tests

The Test Compiler automatically generates either one test or a set of tests. The test or tests are given a name by the lecturer. The lecturer then defines which students should do the test, and when and where they can attempt it. If there is a set of tests then the number of attempts should be specified. This approach is very efficient for lecturers. It allows students to have multiple attempts at the test, thus improving their skill and understanding, and is generally very cheap to administer as there is no need to produce the paperwork, schedule rooms for the tests, organise the examination sessions and staff to invigilate, invigilate, mark the scripts etc. A full automation of the movement of marks is possible but we still like to look at the marks etc. before handing them to the university.

The TAL system is used in three modes. When the number of students and the number of questions on the Main Topic is small, all the students take the test together. Larger classes can be split and each group given a different time to do the test and a different but equivalent test. This second mode does not require very large numbers of questions on the relevant topics. For large groups (320students), when we have a large number of relevant questions, we allow the students to do the test at any time between 0830hrs and 1730hrs on weekdays during two consecutive weeks in the main computer laboratories. We schedule one time when we guarantee access to the PC-lab for an hour. For the mathematics progress tests, students are allowed to take the test three times (it will be a different set of questions each time) and their best score counts. This encourages them to practice their mathematics and has been shown ($p = 0.05$) to benefit their end of year exam results. Students are required to do their working in a bright yellow covered notebook while they are doing a test. In this way we, and they, can look back at their working to see the mistakes they make and potentially check on students whose results are suspicious. At the end of the test, students can review their whole test and are encouraged to copy the feedback for questions they got wrong, into their yellow notebook.

Students Feedback

The results displayed below are from the group of 320 students who were tested on mathematics in their first year using the Open Book, largely Open Access method described above. The questionnaire was distributed and analysed by a final year student. Not all students filled in the questionnaire, nor did they all answer the same number of questions.

Feedback from Student Questionnaire

			Number of responses for each score					
No	***Question***		**1**	**2**	**3**	**4**	**5**	
1	How do you rate the ease of use of the system	Easy	44	50	23	6	4	difficult
2	I would prefer the progress tests to be:	written	20	15	28	39	26	computer marked
	My reasons are:							
3a	I find doing maths at a computer difficult	yes	10	20	25	41	31	no
3b	I like to be able to have more than one try at the test	yes	78	21	12	5	6	no
	I can sometimes give very quick answers to a question because:							
4a	I have seen the question before	yes	6	18	18	41	44	no
4b	I can guess the answer without working it out	yes	17	19	34	31	16	no
4c	Someone will tell me the right answer	yes	2	2	3	26	94	no
5	An immediate response is something I	like	40	32	31	15	9	dislike
6a	The timing bar helps me	yes	43	36	18	17	14	no
7	The time allocated to the progress tests is too	little	17	51	53	5	2	much
8	As a whole the questions asked in the Progress tests are too	easy	0	4	68	49	7	hard
9	The questions are well stated and clear	yes	7	41	31	38	8	no
10	Feedback provided when I get a question wrong is helpful	yes	10	34	33	37	14	no
11	The computer marked tests are	fair	16	42	27	31	12	unfair
12	I guess the answer to a question	often	2	16	34	59	17	never
13	I study further before trying the test again	often	28	56	30	11	3	never
14	A span of two weeks is sufficient to do the tests	yes	91	25	6	3	3	no
15	I find Open Book tests helpful	yes	62	28	15	14	9	no
16	The workbook is useful	yes	21	32	42	21	11	no
17	A 20% penalty for a wrong answer is	fair	16	18	10	36	48	unfair

As a quick summary of these results we can say that:

- students prefer computer based testing because they are allowed multiple attempts at a test;
- there are not enough questions yet to provide the number of tests required to make every student think he has his own special test;
- many students are encouraged to study more because they can have multiple attempts at the test;
- students don't like penalties for wrong answers, but despite the penalties the average mark is 69%.

Conclusions

Over a period of four years we have gradually improved the TAL system to the point where staff in Bristol can use the system to set tests and administer them as largely Open Book unscheduled tests with an adequately high level of confidence in the results. The reliability of the results has improved each year. The use of the system in this mode saves large amounts of staff effort. The huge investment in the database of questions cannot be justified for

Bristol use only. We are now offering a service to any university who would like to use TAL either just using the database questions or, more likely, adding some of their own questions as well. The number of questions on the database has doubled in the last year and with the help of staff from all over the country the database will expand to provide a huge range of questions on many different subjects.

Acknowledgements

The authors thank the Higher Education Funding Committee JISC Technology Applications Project 2/352 for two years worth of support and the University of Bristol for additional support.