LINKING ON-LINE ASSESSMENT IN MATHEMATICS TO COGNITIVE SKILLS

Jane S Paterson

Linking on-line Assessment in Mathematics to Cognitive Skills

Jane S Paterson
Scottish Centre for Research into On-Line Learning and Assessment
School of Mathematics and Computer Science
Heriot-Watt University
Edinburgh

j.s.paterson@hw.ac.uk

Abstract

The project, undertaken in 2000/2001, investigated the ability of on-line assessment to test the skills in Bloom's cognitive domain (Bloom et al, 1956) in the Scottish Qualifications Authority (SQA, 2002) Advanced Higher in Mathematics. There were three main sources of data.

1. The course information: The SQA learning outcomes, content detail and performance criteria.
2. The assessments: The National Assessment Bank questions (NABs), SQA specimen exam papers and the SCHOLAR project (Paterson, 2001; SCHOLAR 2002) on-line assessments.
3. The computer software and data: The online questions and features of assessment engines.

The intention was to match the cognitive skills required in the learning outcomes with those tested in the questions through examination of the key verbs used in each. It was envisaged that the limitations of using one assessment system could be addressed by focussing on features that were available elsewhere. Many difficulties arose and led to a complete rewrite of the on-line questions with closer reference to the performance criteria, the paper based specimen exams, the features of on-line assessment systems and the marking scheme used.

This paper addresses:

a) the subjective nature of the choice of which cognitive skills are expected to be tested in each learning outcome
b) the use of cognitive skills to identify levels of learning in Mathematics
c) the use of key verbs in Mathematics to identify the cognitive skills
d) the approach taken of trying to convert paper based systems for on-line use
e) the features and limitations of existing on-line assessment in producing summative assessment
f) the need for further development in particular areas to increase the scope of summative on-line assessment

Background

There is an apparent lack of research, particularly in Mathematics, into the effectiveness of computer-based assessment to test cognitive skills such as those clearly defined by Bloom although recently the discussions have gained momentum with Beevers and Paterson (Beevers and Paterson, 2002) offering some theoretical considerations on linking outcomes and assessments and leading a series of discussions papers (Beevers and Paterson, 2001).

Project Structure

The project aimed to identify which of Bloom's cognitive skills were expected in the SQA Advanced Higher in Mathematics and relate them to the online assessments in CUE (2002) acting as practice summative assessment in the SCHOLAR project through the following tasks:

a) Determining from content, learning outcomes and performance criteria, the cognitive skills required in terms of Bloom’s Taxonomy for the Advanced Higher in Mathematics with verification through The National Assessment Bank questions (NABs) and the two exam papers (specimen and 2001)
b) Identifying key verbs associated with these skills and their relevance within Mathematics
c) Examining the question structures for automatic assessment to identify any modifications required to the wording of a question for automatic delivery and marking compared to paper delivery
d) Comparing the features of different question types in assessing the skills
e) Analysing the effectiveness of automatic assessment in testing the skills required

Why Use Bloom's Taxonomy?

Bloom's original taxonomy is well known and understood although there has been little work done using it within the on-line assessment area. The following is an abridged version of the cognitive skills definitions given in McCormick and Pressley (McCormick and Pressley, 1997) and clearly indicates the nature of the skills:

knowledge: The ability to know specific facts, common terms, basic concepts, principles and theories.
comprehension: The ability to understand, to explain and to interpret.
application: The ability to apply facts and concepts to new situations and to solve problems.
analysis: The ability to break down a situation into its component parts, to distinguish between facts and inferences, and to identify the organisational structure of the whole.
synthesis: The ability to integrate many ideas into a solution, a conclusion or a generalisation.
evaluation: The ability to judge the quality of something based on criteria and standards.

The taxonomy has been recently revised (Anderson et al, 2000) and shifts the focus to the action verb rather than the skill level. There is also a switch in the two highest skill levels of **synthesis** and **evaluation**. The new structure is very similar to Schoenfeld's Theory of Mathematical Problem Solving (Schoenfeld, 1985) which in turn closely mirrors Polya's four point plan (Alfeld 1996). Although the changes should make skills analysis easier, the same problems arise over the definitions and meanings that verbs can have. King and Duke-Williams (King and Duke-Williams, 2001) have adopted this new taxonomy in their project on higher order skills but the simpler original taxonomy looked more favourable for this project.

Examination of the Course Content and Key Verbs

Here is an example of an outcome and the related performance criteria:

OUTCOME 3: Understand and use complex numbers:
Performance Criteria –

a) Perform a simple arithmetic operation on two complex numbers of the form a + bi
b) Evaluate the modulus and argument of a complex number
c) Convert from Cartesian to polar form
d) Plot a complex number on an Argand diagram.

The mathematical skills required to complete each performance criteria were identified and related to the cognitive skills levels of Bloom's taxonomy. This was a particularly subjective exercise and verification through examination of the summative SQA assessments with marking schemes was felt necessary. The exercise highlighted that the skills level does depend on the educational level under scrutiny.

In a further effort to consolidate and justify the skills chosen for each task, a list of key verbs related to each cognitive skill was compiled from a cross section of web sites. It was hoped that these verbs would clearly link to the verbs used in the performance criteria, the on-line assessments and the paper exams. From the web sites, the 15 most common verbs identified in each category are:

Knowledge	*Comprehension*	*Application*	*Analysis*	*Synthesis*	*Evaluation*
list	explain	solve	compare	design	judge
name	describe	use	distinguish	create	appraise
define	discuss	demonstrate	analyse	formulate	assess
label	interpret	apply	contrast	plan	compare
identify	summarise	compute	differentiate	compose	support
state	restate	operate	separate	modify	criticise
reproduce	estimate	relate	categorise	organise	evaluate
recall	classify	show	infer	develop	conclude
select	translate	illustrate	discriminate	combine	rate
describe	paraphrase	practice	select	write	justify
recognise	predict	calculate	outline	rearrange	defend
match	rewrite	manipulate	relate	devise	discriminate
repeat	distinguish	employ	break down	construct	contrast
write	convert	predict	question	arrange	select
record	generalise	discover	subdivide	integrate	choose

The full table of key verbs indicated that there is considerable debate over which verbs relate to these skills with many found in several skill sections. It was apparent that the verbs associated with the cognitive skills depend on at least 3 aspects:

1. the educational level of the user. For example the verb 'syllogise' is given by one site under **analysis.**
2. the nationality of the user. Certain verbs are used more frequently in some countries and different perceptions arise of terminology. For example, 'figure out' under **analysis** on an American site is rather informal for a British audience whereas the verb 'abstract' (**synthesis)** used in America is likely to be rephrased in Britain.
3. the academic subject under consideration. Several verbs were particular to one academic subject. Verbs such as 'paraphrase' and 'interact' (both **comprehension**) would rarely if ever be seen in a mathematics examination. Other verbs such as 'differentiate' (**analysis**) and 'evaluate' (**evaluation**) take on an entirely different meaning. McCabe, Heal and White (McCabe, Heal and White, 2001) also comment on this.

A more appropriate list was compiled by using the criteria that the verbs:

1. were generally acceptable as testing the relevant skill (Words such as 'describe' (**knowledge**) and 'discuss' (**comprehension**) which appear to be in the wrong category in Maths terms were removed.)
2. made sense in a mathematics context (Those words that are clearly unusual in a mathematical context at this level, such as 'discriminate' (**analysis**), 'reproduce' (**knowledge**), 'discuss' (**comprehension**) and were removed.)
3. were likely to be found in the assessments at the educational level under scrutiny. (There were some words such as 'appraise' (**evaluation**) that would not be used at this level and were omitted.)

The amended lists are shown:

Knowledge	*Comprehension*	*Application*	*Analysis*	*Synthesis*	*Evaluation*
name	explain	solve	compare	formulate	assess
`define	describe	use	analyse	modify	compare
identify	estimate	demonstrate	differentiate	combine	evaluate
state	classify	apply	select	write	justify
select	rewrite	show	break down	rearrange	select
write	convert	illustrate	identify	devise	choose
show	extend	calculate	calculate	construct	estimate
draw	express	predict	deduce	integrate	interpret
	identify	construct	evaluate	generate	explain
	illustrate	sketch	solve	rewrite	determine
	give	find	estimate	deduce	verify

In many cases it is still possible to argue on precisely which skill is tested in Mathematics by a particular verb and this is apparent with the inclusion for example, of 'calculate' under **application** and **analysis** or 'select' under **knowledge, analysis** and **evaluation.**

Similar difficulties arose on comparison of these web site key verbs and those used in SQA summative assessments. Several key verbs used in the tests were not contained in the keyword lists: for example, 'plot', and 'prove'. Other key verbs needed to be considered in context (for example, 'determine', 'write' and 'differentiate').

The investigation revealed that most of the key verbs used fell into one of the first three skill categories, contradicting several of the perceived skills determined from the performance criteria. The conclusion was that key verbs were not as appropriate to classifying cognitive skills levels in Mathematics as they may be in other subjects.

Analysis of the Questions

ORIGINAL QUESTIONS: The original questions set for practice summative assessment caused too many difficulties to be helpful. In practice summative assessment (help mode), the questions were designed to give as much assistance as possible to the less able student. The features of help mode include:

- Steps: these were included to break down the part(s) of a question in such a way that it helped the weaker student to progress through the question. The skills tested through steps were therefore of a lower cognitive level than the questions set within the key parts and the level could not go beyond **application**.

- Keyparts: these had often been used to break down the question in a similar manner to the steps. This is a feature of automatic assessment that is of general concern and has been highlighted recently by Lawson (Lawson, 2001). He also points out that the breakdown of a question can in fact introduce extra learning outcomes but examination of the types of extra outcomes suggests that these may be at a lower cognitive skill level. Although the effect of keyparts was less obvious than steps, it still compromised the skills level of the questions by forcing a method on the user or providing more information.

- The ability to reveal an answer: this feature also suggested a lowering of the skills level by providing additional information.

- The informal wording of the questions: this caused concern. In Mathematics, questions such as ‘What is…?’ are common. Other questions would ask ‘Give the coordinate…’ but the skills required to do so were far more complex than the question would suggest.

Additionally the verbs relating to Bloom’s cognitive domain were chosen using the precise definition of the verb and did not take into account the context in which the verb was being used. An obvious set of examples where problems occur in mathematical questions are: 'differentiate', 'integrate', 'show', 'give', 'determine', 'evaluate', and 'use'. Aspects of the exam papers such as marks awarded for method had also not been addressed in the questions.

NEW QUESTIONS: The rewritten questions concentrated on trying to provide questions closely related to the full range of skills required within the particular performance criteria. The marking scheme played a prominent part in the setting of these new questions and frequently helped to identify the required skills and the likely structure of the questions.

Matching New Questions to the Marking Scheme

A major problem centred on matrices. Although it is possible to use matrices (and thereby vectors in component form) within the question, there is no provision as yet for the marking of a full matrix, a determinant or a vector in symmetric form in any engine.

There were also problems with number theory proofs. Although McCabe, Heal and White try to provide a solution by using *drag and drop/matching*, the method is not generally accepted as being equivalent to the answer constructed from scratch on paper.

There were difficulties with the constant when integrating. The paper version does not deduct marks if it is missing but there is no way of marking multiple correct solutions in the *algebraic* question type.

There were many questions where the method of reaching a solution gained marks. In these cases, it can be argued that when the method is being tested, this, in effect, is testing the **application** skill. Without *adaptive* testing, every question had to be worded in a clearly different manner to the paper version. Two examples are the method of obtaining partial fractions and finding an integral where integration by parts is needed.

Synthesis skills were normally required in questions asking for a final completely simplified solution. This again required the on-line questions to be worded very differently and in such a way that the user is guided to the solution.

Do Question Types Perform Similarly?

Although terminology differs between software developers the following list gives the base types of question that may be offered:

multiple choice:
multiple response:
algebraic expression:
numeric:
word match (or text):
hot spot:
gap fill:
drag and drop:
graphing:
ordering:
essay:

(*Essay* marking is being explored in America using norm-referenced marking but there is much work to do before *essay* style will be automatically marked using criterion referencing. It cannot be considered as a complete on-line assessable question type at present.)

There is a great deal of debate over what types of question are suitable for particular skills. McCabe, Heal and White suggest that *algebraic*, *numeric*, *text*, *multiple-choice* and *multiple response* only test foundation thinking but that *hotspot*, *drag and drop*, *ordering* and *essay* test higher order skills. Neither the author nor Lawson support this view.

The project itself indicated the following use of question types:

- *Multiple choice* or *multiple response* in Mathematics is problematic. It offers a choice to the user and thereby provides information which the user would not have taking a paper test. It can be used as a quick test of **knowledge** but is perhaps not useful in summative testing of Mathematics unless all possible answers can be included in the choices.

- Of prime importance is *algebraic* expression. It extends the possibilities for testing and can push the skill level as high as **synthesis** making it more closely aligned to the paper based questions. However, some improvements are required to cope with the breakdown of marks where method in, for example, **application** skill tasks, is important but not unique.

- *Numeric* question types offer a quick convenient way to input answers and have a surprising range of settings. The range of skills tested varies from question to question.

- *Word match* has a limited use at the **knowledge** level in identifying, for example, maximum or minimum turning points or types of proof. Its use within Maths questions may extend to **comprehension** skill level but no higher. Once *word match* is developed further to include comprehensive *phrase matching* and ultimately *essay*, this type will begin to be useful in proof marking.

- *Hotspot* can be useful for identifying graphs of functions or maxima etc but in this context is little more than *multiple choice*. It does however provide for questions that cannot be set in any other way. It tests **knowledge** through recognition on diagrams.

- *Gap fill* is very powerful for languages but has little use in Mathematics until Maths expressions can be included and recognised in equivalent mathematical forms.

- *Drag and drop* can be used in graph drawing (Mathwise 1998) or in compiling proofs (McCabe, Heal and White 2001) but, as with multiple choice, all the necessary information is given thus reducing or removing some skill level.

- *Graphing* (Triads 2001)) is particularly useful for simple graph drawing but at Advanced Higher level, the sketches required are too complex to be done in this manner.

An overriding problem with the on-line testing of **application**, **analysis** and **synthesis** skills is the existing practice on paper of awarding marks for the process of reaching the answer, for working and for correct techniques. In on-line assessment this scheme looks doubtful without lessening the skills level until other question types such as *hidden multiple choice, equation, essay*, *follow through* and *adaptive* testing are operational. There is too little information to make a judgement on **evaluation** but given the type of question likely to be set, it seems unlikely that it can be tested by on-line assessment.

Conclusions

The conclusions confirm the widely held belief that automatic assessment can test the two lowest cognitive skills but cast doubt on the ability of automatic assessment to soundly test the remaining four levels.

It is not appropriate to take paper-based assessments and try to format them for automatic delivery. Consideration must focus on the skills that should be tested automatically and how these can be related precisely to tasks. This may lead to the restructuring of learning outcomes, question wording and presentation in testing skills without the bias of paper assessment comparison.

This research depended greatly on the ability of the researcher and others to determine the skills required for any task. Although attempts were made to verify the skills decisions reached, in general there does not appear to be any definitive method for doing this. The research already published supports this by failing to offer any reasons for the skills choices made (McCabe, Heal and White, 2001; King and Duke-Williams, 2001). The author accepts that this is an area that requires further research work.

Care should be taken with automatic assessment in ensuring that the structure of the question does not change the skills that it tests (or introduce new ones) (Lawson, 2001).

At present no assessment engine provides all question types and this makes it difficult to state which skills can be tested overall. Continuing development is needed to remove the problems with missing types.

An overriding concern throughout the project centred on justifying which skill related to a mathematical task. This project highlights the problems of using key verbs and suggests that consideration should be given to alternative approaches. One such approach would be to concentrate on the way in which the student thinks when performing a task and this may lead to research into the role of concept maps/mind maps and strategic thinking in assessment of students.

References

Alfeld, P. (1996) *Summary taken from G. Polya "How to Solve It" (1957) 2nd edition.* Princeton University Press
http://www.math.utah.edu/~alfeld/math/polya.html (22 Apr 2002)

Anderson, L. W., Krathwohl, D. R. and Bloom, B. S. (2000) *Taxonomy for Learning, Teaching and Assessing, A: A Revision of Bloom's Taxonomy of Educational Objectives, Complete Edition,* London, Longman,

Beevers, C. E. and Paterson, J. S. (2002) *Assessment in Mathematics*, **in** Kahn, P. and Kyle, J. (eds) Effective Learning and Teaching in Mathematics and its Applications, London, Kogan Page

Beevers, C. E. and Paterson, J. S. (2001) *Automatic Assessment of Problem Solving Skills in Mathematics,* Paper presented at International Conference on Communication, Problem Solving and Learning, Strathclyde University. To appear 2003. Active Learning in Higher Education, ILT, London, Sage Publications
http://ltsn.mathstore.ac.uk/articles/maths-caa-series July (22 Apr 2002)

Bloom, B. S. Engelhart, M. D. Furst, E. J. Hill W. H. and Krathwohl, D. R. (1956) *Taxonomy of Educational Objectives, The Classification of Educational Goals. Handbook 1: Cognitive Domain,* New York, Longmans

CUE Assessment system, (2002) Heriot-Watt University
http://www.scrolla.hw.ac.uk (22 Apr 2002)

King, T. and Duke-Williams, E. (2001) *Using Computer-Aided Assessment to Test Higher Level Learning Outcomes*, Proceedings of 5th International Computer Assisted Assessment Conference, Loughborough

Lawson, D. A. (2001) *Computer Aided Assessment in Relation to Learning Outcomes*
http://ltsn.mathstore.ac.uk/articles/maths-caa-series Oct (22 Apr 2002)

Mathwise (998) *Pre-Calculus CD,* Oxford, Numerical Algorithms Group

McCabe, M. Heal, A. and White, A. (2001) *Computer Assisted Assessment (CAA) of Proof = Proof of CAA New Approaches to Computer Assisted Assessment for Higher Level Learning,* to appear in Proceedings of 5th International Conference on Technology in Mathematics Teaching, Klagenfurt, Austria
http://ltsn.mathstore.ac.uk/articles/maths-caa-series Sept (22 Apr 2002)

McCormick, C. and Pressley M. (1997) *Educational Psychology, Learning, Instruction, Assessment,* New York, Longman

Paterson, J. S. (2001) *The SCHOLAR Project*. Academic Exchange Quarterly Journal, Winter 2001/2002, New York

SCHOLAR project, (2002) Heriot-Watt University:
http://www.scholar.hw.ac.uk (22 Apr 2002)

Schoenfeld, A. (1985) *Mathematical Problem Solving*,
http://www.gwu.edu/~tip/schoen.html (22 Apr 2002)

SQA (Scottish Qualifications Authority), (2002)
http://www.sqa.org.uk (22 Apr 2002)

TRIADS toolkit, (2002) Centre for Interactive Assessment Development
http://www.derby.ac.uk/assess/newdemo/newdemo.html (22 Apr 2002)