ASSESSING HIGHER ORDER SKILLS USING SIMULATIONS

Ruth Thomas, Helen Ashton, Bill Austin, Cliff Beevers, Dave Edwards & Colin Milligan

Assessing Higher Order Skills Using Simulations

Ruth Thomas[1], Helen Ashton[1], Bill Austin[2], Cliff Beevers[1], Dave Edwards[1] and Colin Milligan[1].

1. School of Mathematical and Computing Sciences, Heriot-Watt University.
2. JeLSIM Partnership. www.jelsim.org

Abstract

The alignment of assessment strategy with learning outcomes is important as it ensures the validity of any measurement obtained. Without the ability to assess higher order skills, online assessment will be unable to progress beyond objective testing. Bennett's (1998) vision of the future of assessment sees an increased use of simulations together with a blurring of the lines between assessment and teaching. This paper outlines a project that is taking the first steps in this direction.

A system is being devised to allow an assessment engine and a simulation to communicate at a deep level to improve authoring capability and enable more complex assessments. The paper explores issues that have been addressed in the design of the communication interface and protocols. The higher order skills that can be assessed with the system are outlined with examples, using the cognitive process dimension of the revised version of Bloom's taxonomy produced by Anderson et al (2001)

Introduction

Quality of assessment is central to the quality of learning. As Ramsden (1992) points out: *"The process of assessment influences the quality of student learning in two crucial ways: it affects their approach and, if it fails to test understanding, it simultaneously permits them to pass courses while retaining the conceptions of subject matter that the teachers wishes to change"*

The importance of aligning overall assessment strategy, individual assessment methods and the criteria used in judging the quality and standards of learning with the teaching and curriculum objectives and intended learning outcomes is being increasingly recognised (Anderson et al (2001)).

Constructivist learning techniques in which active and self-directed learners acquire knowledge and skills undertaking authentic complex tasks cannot be easily aligned with traditional assessment techniques (Elshout-Mohr et al (2002)). Ultimately, skills that have been learnt will be applied in the real world. Consequently, authentic assessment that is aligned to the requirements of the specific environment in which the student will apply knowledge is important if a student's true potential is to be measured. Assessment strategies must change in order to meet new needs.

How then does automated, online assessment fit within the push for higher quality? One perception is that the search for efficiency via automated mechanisms may work against the drive to increase quality. With its over reliance on objective-testing, online assessment is seen by many as useful for assessing lower order skills such as recall of knowledge, whilst being ill-equipped to assess higher order skills such as the ability to apply knowledge in new situations or to evaluate and synthesise information. This need not be the case, as long ago as 1956, Bloom et al (1956) modelled multiple choice item formats for categories in their taxonomy demonstrating that they could be used at all levels. Even so, Anderson et al (2001) point out that despite technological advances there has been a distinct lack of progress in such formats over the past 40 years.

Bennett (1998) suggests that online assessment has not yet achieved its full potential. He states, *"Like many innovations in their early stages, today's computerised tests automate an existing process without reconceptualising it to realise dramatic improvements that the innovation could allow".* He sees three stages in the growth of online assessment. In the first stage, tests resemble paper-based tests, though there is use of adaptive technology. In stage two there is an increase in the use of new formats such as multimedia and constructed responses and in the final stage, much greater use of complex simulations and virtual reality will occur along with a seamless embedding of assessment within learning.

New formats and responses have been used in an effort to assess higher order skills. In some subjects there have been attempts at using free text input to extend automated assessment beyond multiple-choice questions. Commercial offerings are beginning to appear e.g. from Intelligent Assessment (Mitchell et al, 2002) and e-Rater from the Educational Testing Service (http://www.ets.org/erater/). In the mathematical domain, work has been undertaken to test higher order skills using a computer algebra system (CAS) to check student answers in the background (Sangwin 2003).

Simulations[1] can be used to provide activities that support education where learning outcomes require more than the acquisition of knowledge. They bring both reality and interactivity to eLearning, allowing learners to manipulate a system directly and to observe the effect of the change, thus providing a form

---

[1] Simulations are defined as a model of a system that can be explored by the learner, changing input variables results in a change in output variables

of feedback that facilitates exploration, allowing learners to build their own understanding. To date, simulation use in automated assessment has been limited. Bespoke assessment systems involving either hardware or software simulators exist (e.g. Lapointe et al (2000)) and in some cases, human assessors evaluate the student's performance on a simulator (e.g. Cleave-Hogg (2002)). TheTRIADS assessment system (MacKenzie, (1999)) has a variety of question styles some of which are equivalent to simple simulations.

What does not currently exist is a system that can make use of pre–existing simulations or other applications in the assessment process. The aim of this work is to report on a project whose ultimate intention is to do just this.

System overview

In general, online assessment systems consist of the following components:
1. Authoring tool(s) to allow non-programmers to create questions and tests.
2. A delivery engine to present these questions to the user in the appropriate manner.
3. An events database/log which allows review of progress, question design etc.
4. Other components such as administration tools to generate and manage logins and test availability.

Work is being undertaken to integrate simulations in the first three of these components. This was not just a case of allowing the simulation to exist within the system (something which has been possible for a number of years), but allowing the assessment system and the simulation to communicate at a much deeper level to improve authoring capabilities and enable more complex assessments which can be managed, assessed and reported in an appropriate manner.

Potentially, the techniques used could enable any interoperating process to communicate any information between itself and the assessment engine. The first phase of the project is focussed on the interoperation of simulations with the assessment engines. Subsequent phases will look at interoperation with external applications other than simulations.

The potential

The integrated simulations can be used in three ways:
• As part of the assessment question;
• To provide answer mechanism /new response forms (when using a simulation, the type of response that a learner can make to a question can be any activity that can be carried out within the simulation, e.g. sketching a line on a graph, building a model, carrying out a sensitivity analysis);
• To provide feedback to the learner.

Utilising simulations in this way provides opportunities for increasing both the quality and efficiency of assessment. As mentioned previously, simulations are a powerful educational resource: the ability to link them with an assessment engine will allow students to be assessed in the same environment in which they learn and will integrate the assessment activity more fully into the educational process. The forms of assessment could be diagnostic or formative assessment providing guidance and help as the student carries out a task or summative assessment of an entire process providing a measure of competency at fulfilling a task. Rather than asking questions of the form "How would you…" it would be possible to follow students as they undertake a process and monitor the choices involved. Thus, assessment can be closely aligned to learning outcomes, even when those outcomes involve higher order skills. The final section expands on this in more detail. The next section considers the functionality of the system.

Functionality

The first stage of the project involves the production of a prototype of the system working with the PASS-IT Assessment Engine (PAE, previously CUE (Paterson et al 2002) and the JeLSIM toolset[2]. (Thomas and Milligan, (2004)). Although the current focus is on communication with the JeLSIM toolset, the techniques also allow PAE to intercommunicate with Flash (Macromedia) and ultimately any external media. Use of the JeLSIM toolset has an advantage over other technologies, because it allows non-programmers to use click and drag techniques to construct or modify the simulation visualisations used in setting questions.

A major issue in building such a system has been the specification of the communication between components, i.e. what is communicated and when. A number of these issues have been addressed in the submission of use cases to the IMS Global learning Consortium (Milligan et al (2003)). An important goal for the system has been that once a simulation has been integrated into the assessment engine, it should not be necessary to modify the simulation model (program code) in order to re-use it to produce a range of different question types. In the prototype system, the following key issues have been addressed in the design of the communications interface.

1. **The initial state of the simulation**. This can be controlled from the assessment engine to allow different question scenarios to be set using the same simulation model. By starting a simulation in a different state, it is possible to set case studies or to ensure that the simulation is demonstrating a specific phenomenon relevant to the question.

2. **Randomisation of simulation state:** PAE provides the ability to randomise variables: this has been incorporated into the system so that the question setter can choose which of the simulation variables can be

---

[2] JeLSIM software separates the behavioural model of the simulation from the visualisation. One model can be used to produce many visualisations. Simulations can be easily deployed to the web as Java applets.

randomised and over what range. This is different to the issue raised in item 1 as is used to provide alternative but equivalent starting conditions (not different scenarios). In using this capability, care must be taken that the simulation is not placed in an anomalous state or error condition.

3. **Where should the "accuracy" of the student activity be assessed?** Theoretically this could be done within the simulation or within the assessment engine. For simulations involving complex calculations it is pointless to duplicate that calculation at the assessment engine so only simple arithmetic checks will be made there (e.g. is variable $X > Y$, does variable K fall within a specified range).

4. **Where should the question be marked?** Once the accuracy of the answer is assessed, in keeping with the overall goal of minimising alteration to the simulation in this process, marking remains the domain of the assessment engine.

5. **Where should the feedback come from, the simulation or the assessment engine?** Simulations provide intrinsic feedback as the learner uses them; they sometimes also have generic feedback in the form of error and warning messages. Feedback which tells the learner the correct answer and whether their answer was right or wrong is most easily provided by the assessment engine, feedback given to the learner about why they are wrong or which seeks to guide the learner is best given in the simulation. Once the feedback has been given, it is useful to communicate it to the assessment engine for storage as part of the reporting functionality to aid reflection in both learner and teacher (Ashton et al (2004)).

6. *Preservation of state:* After a student has modified a simulation, the state should be preserved so that a student returning to that question, or refreshing the web page during a session, will always be able to view, and then modify the current state of the simulation.

7. *Passing "answers" from the simulation to the assessment engine.* . The question setter can add a "submit" button to the simulation component of the question. When pressed this passes out a value to the assessment engine. Given the great range of simulations and question types that might be produced by the system, great flexibility is required in terms of what is passed out and the question setter can choose to pass out a single value, or an expression involving variables. E.g. $X - Y / Z$. It is also possible to pass out arrays or elements of arrays.

8. *Questions with parts and subparts:* PAE allows questions to be divided into different key parts. This allows a series of related questions to be asked from a single simulation. Submit buttons added to the simulation can be tailored to either send information about a single key part or a number of selected key parts to the assessment engine.

9. **Reporting database:** PAE has a powerful reporting database (Ashton et al (2004)). The aim is to record information so that the data is useful to both students and teachers. By saving information about changes student made to simulation variables, information would be available about how students handled an activity, for instance did they randomly guess, or experiment in a controlled manner.

10. **Marking schemes**: PAE provides the capability of flexible marking, including partial credit. This combined with its delivery of a single question as a number of key parts means that flexible, individualised marking scheme could be devised for any simulation based activity assessed by the engine.

11. **Authentic tasks**: In testing higher order skills, it is important that students are able to undertake meaningful tasks where a series of decisions/actions are undertaken. The assessment process should not be overly intrusive; students should not be required to explicitly send each decision made to the assessment engine whilst using the simulation. Tasks may have a series of stages at which it is useful provide feedback before allowing the student to proceed. Use of key parts and customisable submit buttons is designed to provide the necessary flexibility.

Potential Usage

In this section, the way in which the system might be used to assess higher order skills is outlined. The section is structured using the higher elements of the cognitive process dimension of the revision of Bloom's taxonomy (Bloom et al 1956) undertaken by Anderson et al (2001). The processes, in order of education level, are *remember, understand, apply, analyse, evaluate and create*. A number of possible examples are given for each process; some are being developed as exemplars using the prototype system and will be made available online at http://calm.ac.uk/higherOrder.html. N.B. They are intended to demonstrate the principles and promote ideas and generate feedback, they do not attempt to provide the perfect question.

Assessing Understanding

• **Exemplify:** (Find specific examples or illustration of a concept or principle) In a simulation context, this could involve: driving the simulation to a state that shows a phenomenon in action, (e.g. resonance in a simulation of vibrations); or satisfying some pre-existing criteria (e.g. break even in a business simulation). The exemplar question involves the simulation of the polarization of light passing between two media, the student is asked to set the angle of incident to the Brewster angle.

• **Predict:** (Draw a logical conclusion from presented information) In terms of simulation use the learner might be asked, "What will happen next?" The exemplar question uses the simulation of a chemical reaction and asks, "What will happen if the temperature is increased?" The learner can

mark the expected change on a graph. The simulation provides feedback of what actually happens.

- **Compare:** (Detect correspondences between two ideas, objects etc)
In the simulated domain, this could involve mapping one system into an analogous system, (e.g. using an analogy between electricity and water). It could also cover the mapping of a real world system into a computer model of a system, for example asking the learner to model a vibrating violin string or bridge as a series of masses, springs and dampers in a generic simulation of vibrating systems. The exemplar shows a question that states a real world scenario and asks the student to map it into the parameters used within a simulation of projectile motion.
- **Explain:** (Construct a cause and effect model of a system)
In assessing activities that require the learner to provide an explanation, use of free text entry in combination with the simulation could be a useful pairing of techniques. The provision of explanation should not be seen as an isolated activity, the "Prediction-Observation-Explanation" (POE) technique is a powerful way of challenging learner's alternative conceptions (e.g. White and Gunstone (1992)) (Jimoyiannis (2001)). The exemplar shows a mixture of simulation and multiple choice questions to demonstrate an extension of the "compare" type question to produce a POE question

Assessing Application

- **Execute** (Apply a procedure in a familiar task)
In a simulation context, this might involve routine application of a procedure to a set of data. Typical examples would be: use of a dataset to calculate the age of a rock using radioactive dating; or calculation of the rate of a simulated chemical reaction.

- **Implement** (Apply a procedure in an unfamiliar task)
In order for a task be classified as implementation rather than execution, it must require the learner to invoke conceptual knowledge as well as knowledge of the technique. Instances of this would be: a simulated case study where the learner must decide whether a business is solvent; or questions involving an experimental dataset that is subject to experimental error.

The exemplar demonstrates the assessment of "application" by showing the way in which a longer task (the determination of activation energy of a reaction using the Arrenhuis equation) might be assessed.

Assessing Analysis

- **Differentiate:** (Distinguish relevant from irrelevant)
A typical assessment of analysis would be to provide a simulated case study and ask the learner to determine the important factors contributing to the current state. For example, asking the learner to identify the key factors leading to a decline of a population using a population dynamics simulator. The exemplar consists of a simulation of coulombs law in which the charge on one of the particles is unknown and students must manipulate the system, ignoring irrelevant information to determine it.

- **Find coherence:** (Determine how elements fit or function within a structure).

In terms of a simulation, this type of skill covers exploratory activities that allow the learner to build a feel for the relationship between the underlying factors governing a system and promote an appreciation of appropriate ranges for system parameters. In the exemplar, the learner is provided with a system allowing them to explore the factors affecting reaction rate. Their discoveries are assessed using multiple choice questions (though note that free text answers would be the ideal answer format in this case.

Assessing Evaluation

- **Check:** (Detect inconsistencies or fallacies within a process or product)

Assessment using simulations in this process could include: trouble shooting simulated systems with broken or malfunctioning components, (e.g. within an electrical circuit); checking if data confirms a hypothesis; and checking if a product meets a design specification. An interesting advance in these cases will be being able to use records of activity stored in the reporting database to see the strategy adopted by the learner in carrying out the checking process.
.

- **Critique:** (Detect inconsistencies, determine appropriateness of a procedure)

Within a simulated domain this might take the form of a case study where it is necessary to judge the merits of a number of solutions with respect to given constraints (e.g. running a national economy).

In the exemplar demonstrating the assessment of "evaluation", the learner is given cashflow and trading forecasts for a small business made using certain assumptions. The learner must consider the viability of the business and look for inconsistencies in the plans. Free text answers would be the ideal, but currently the assessment is carried out using multiple choice/ response.

Assessing Creation

The creation process can be split into three closely linked stages: generating solutions, evaluating and selecting the appropriate strategy, planning and then undertaking an activity that solves the problem. Each stage could be assessed separately, however, the example given below covers use of a simulated laboratory to allow assessment of the creative process involved in the scientific method.

In the first stage, learners generate hypotheses. In a simulated domain, this could involve asking the student to generate a hypothesis to explain a phenomenon, or to come up with a range of hypotheses for a given problem. For example: what are the factors might affect the rate of a chemical reaction? In the next stage, learners design experiments to test the hypotheses. For example: how can the effect of sunlight, particle size or temperature on

reaction rate be measured in a controlled experiment? In the final stage, learners act on the plan from the previous stage and carry out an experiment or design and drawing conclusions.

This style of computerised assessment does more than automate an existing process. The traditional laboratory exercises, usually only test execution skills because the student is usually told the aim of the experiment and a detailed description of the techniques to use is provided.

Conclusion and next steps

The ability to assess higher order skills with simulation will improve the alignment of learning outcomes with assessment. Some of the activities that could be assessed using this system cannot be assessed using traditional techniques.

The prototype system is currently being developed. We are currently exploring the range of questions that can be authored within the system. Small scale pilots of the system will be carried out using the questions in diagnostic, formative and summative assessment. There are a number of questions these pilots must answer. How usable will students find the new style of question, will unfamiliar user interfaces effect performance, will they be acceptable to teachers? For summative assessment, what are appropriate marking schemes and how can we ensure the validity of such assessments?

Comment [RCT1]: Page: 1
Suggestions from those of you who are practitioners in this area would be welcome

An important area to pursue is the integration of assessment into teaching, the ultimate aim being non intrusive assessment and the student being assessed in the same environment in which they learn. Early prototypes of assessment of meaningful tasks of longer duration are being developed using the system. It would be useful, once a student difficulty has been detected to provide some form of assistance, either in the form of feedback personalised to student need or as coaching/ scaffolding to suggest activities that may help the student.

New skills will be required by those authoring questions to take advantage of the system. The plan is to provide guidance to question authors through the use of simulation/question templates that provide a starting point for the construction of a range of common combinations. The aim will be to minimise the time and cost of production, whilst maximising the quality of assessment questions.

References:

Anderson, L. W., Krathwold, D. R., Airasian, P.W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J. and Wittrock, M. C. (eds.) (2001). **A taxonomy**

for Learning, Teaching and Assessing: A revision of Blooms Taxonomy of Educational Objectives, Addison Wesley Longman, New York.

Ashton, H. S., Beevers, C. E., Schofield, D. H. and Youngson, M. A. (2004). **Informative reports - experiences from the PASS-IT project**. To be presented at the 8th International CAA Conference, Loughborough.

Bennett, R. E. (1998). **Reinventing Assessment: Speculations on the Future of Large Scale Educational Testing**, Educational Testing Service. (http://www.ets.org/research/pic/bennett.html)

Bloom, B. S. and Krathwold, D. R (1956) **Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain**.New York; Toronto: Longmans.

Cleave-Hogg, D., Morgan, P. and Guest, C. (2002), **Evaluation of Medical Students' Performance in Anaesthesia using a CAE Med-Link Simulator System**, Proceedings of 4th International Computer-assisted Assessment Conference, Loughborough University, UK.

Elshout-Mohr, M., Oostdam, P. and Overmaat, M (2002)"**Student assessment within the context of constructivist educational settings**." Studies in Educational Evaluation 28: 369-390.

Jimoyiannis, A. and Komis, V. (2001)"**Computer simulations in physics teaching and learning: a case study on students' understanding of trajectory motion**." Computers and education 36: 183-204.

Lapointe, J. and Robert, J. (2000). "**Using VR for effective training of forestry machine operators**." Education and Information Technologies 5(4): 237-250.

Mackenzie, D. (1999). **Recent developments in the Tripartite Interactive Assessment Delivery System (TRIADS).** Third Annual Computer Assisted Assessment Conference, Loughborough.

Milligan, C., Currier, S. and Cross, R. (2003). **SCR-01 Assessing With Active Content, CETIS Use Case for IMS Interactive Content SIG**. (http://www.imsglobal.org/usecases/ic/cetisusecase004_v4.pdf)

Mitchell, T., Russell, T., Broomhead, P., Aldridge, N., (2002) **Towards Robust Computerised Marking of Free-Text Responses**, Proceedings of the 6th International Computer Assisted Assessment Conference, Loughborough.

Paterson, J. S. (2002). **The CUE Assessment System**, LTSN MathSTOR, Maths-CAA Series.

Ramsden, P. (1992). **Learning to Teach in Higher Education**, London: Routledge.

Sangwin, C. J. (2003) **Assessing higher mathematical skills using computer algebra marking through AIM**, Proceedings of the Engineering Mathematics and Applications Conference (EMAC03), Sydney, Australia.

Thomas, R. and Milligan, C. (2004). "**Putting Teachers In The Loop: Tools For Creating and Customising Simulations**." Journal of Interactive Media in Education (submitted).

White, R., Gunstone, R. (1992). **Probing Understanding**, The Falmer Press, London.