

Simulation as a Tool for Computer Assisted Formative Assessment: First Aid as a Case Study

Andrew Young and Steven Cafferty
School of Sciences
University of Salford
Manchester
M5 4WT
Contact email: A.J.Young@salford.ac.uk

Abstract

It is widely recognised that, while formative assessment is good for learning, it is expensive in tutor time. This is particularly true in environments with increasingly large class sizes and with increasing student diversity due to widening participation. Computer Assisted Formative Assessment can help, by providing the opportunity for self-tests without tutor involvement. Often this takes the form of multiple choice questions (MCQs) that allow students to gauge their own progress. However, this is only allowing the students to assess themselves summatively and does not address the learning and feedback issues required for assessment to be truly formative.

We believe that computer-based simulation provides a tool for true Computer Assisted Formative Assessment that has received insufficient emphasis in educational literature. It can provide unlimited formative assessment without tutor involvement, though only if the class of problem is amenable to simulation.

In this paper, we will define what we hope to achieve by Computer Assisted Formative Assessment, we will describe our current ideas, and we will demonstrate them using two First Aid tools as non-academic illustrative examples (a non-academic example has been chosen to allow pedagogic issues to be separated from subject-specific issues).

Finally, we will discuss the work that still needs to be done to make the use of simulation more rigorous and to define the class of problem amenable to such a solution.

We conclude that simulation has a lot to offer the academic community in providing efficient, scaleable and formative Computer Assisted Assessment that complements methods such as MCQs and that can be applied to a wide range of subjects.

Keywords: formative assessment, simulation, equity, validity

Motivation

With increasing class sizes and diminishing resources, the need for increased efficiency in assessment is clear. Computer Assisted Assessment (CAA) is one of the tools in the academic's toolkit to help with this. However, it is not without its problems.

Firstly, the techniques that provide assistance in marking (summative assessment) often rely on techniques that may find it hard to assess knowledge in Bloom's "Evaluation" level (Bloom et al, 1956) or at the SOLO "Extended abstract" level (Biggs, 1999). The consequence of this is that, where this knowledge must be assessed, CAA can only assist in the logistics of the assessment.

The second problem is related. If CAA finds it hard to provide summative assessment of "deep learning" then it will find it equally hard to provide formative assessment. The difficulty is compounded by the fact that students need (and want) vast amounts of formative feedback in order to direct their learning. Indeed, Biggs describes formative assessment as "that synonym of good learning" (Biggs, 1999).

In a normal teaching environment, providing effective formative assessment is very difficult. In Computer Science at Salford, we often teach "small group" laboratory sessions with groups of 30. Given time to explain what the tutor wants them to do, and to wrap up at the end, this allows an average of a minute per student (and the ones who get the "early minutes" haven't yet done anything to give feedback on). It is clear most students get no useful formative feedback at all.

Use of CAA would seem to have advantages here if it could allow the student to take charge of their own formative assessment, and to do it in their own time. However, it's not that simple:

- if the CAA method requires academic moderation (i.e. marking) then the resource implications mean CAA has moved the problem rather than alleviated it; and
- if the CAA will automatically generate the assessment results then it is likely to rely on techniques such as multiple choice questions (MCQ). These techniques aren't ideal for three reasons:
 - it is difficult to generate questions that test "deep learning", as noted by many authors (Race, 2001; Biggs, 1999);
 - there is often very poor constructive alignment (Biggs, 1999) between the learning outcome and the assessment task; and
 - there is the problem of "familiarity" if students see the same problems over again, which can end up testing recall of the answer rather than understanding of the problem (this is not usually a problem for summative assessment, as students only see a problem once; but with formative assessment they may seek assessment many times).

We can summarise this discussion by stating an ambitious objective that defines what we require from Computer Assisted Formative Assessment (CAFA):

“To allow students to give themselves unlimited formative assessment without the involvement of an academic”

This paper describes our current work towards achieving this. It is not a description of a completed system, but is a description and demonstration of the ideas that we have had so far, and some thoughts about what we plan to do in the future.

Approach

In this section, we will define exactly what we mean by formative assessment, and exactly what a CAA system needs to do to support this.

It is our experience that formative assessment is often viewed informally as “the same as summative assessment, but the mark doesn’t count” or “giving the student some feedback on something they’ve done”. We believe the former is not helpful to the student, and the latter is not helpful to the understanding of the pedagogical processes involved, especially where the “feedback” is just “the correct answer” (Davies 2001). We need a more formal approach.

If we use Kolb’s “experiential learning cycle” (Kolb 1984) as a model of learning then we think of a four stage cycle comprising experience, reflection, abstraction and experimentation. If we can encourage (or, ideally, force) students to go round this cycle then we can help them to learn. A precise definition of formative assessment therefore has two components: firstly it *generates* experiences that require the student to reflect and undertake a learning cycle; and secondly it gives the student some feedback about *the way they went round* the learning cycle.

For CAFA, the *experiences* would be computer generated. However, to meet our objective of “without the involvement of an academic”, we also require the *reflection, abstraction* and *experimentation* are computer managed. If the student is asked for some input after each experience then there are three ways that we can handle the input:

- give the student a “model answer” straight away – i.e. tell them what their input should have been and let them self-assess their own performance;
- give the student a summary of their performance at the end – the student is effectively summatively assessing themselves, with little feedback; and
- never give an answer, but build the student’s reflections into subsequent experiences, so that the experiences build on a student’s knowledge.

We believe the first two options encourage the student to reflect only on a single issue at a time (in SOLO terms this is at best “multistructural”) and learning cycles are independent. However, the third option forces students to reflect on the current experience along with their previous reflections (in SOLO terms this can be “relational” or better). Successive learning cycles become dependent.

Consider the following (highly contrived) example. Compare:

Q: What is the moon made of? A: Cheese Q: No, it's made of rock. How heavy was the Apollo 11 lunar lander?
--

with

Q: What is the moon made of? A: Cheese Q: How much cheese would you need to support the weight of the Apollo 11 lunar lander?

The intention is that, in the second case, reflecting on the second question implies also reflecting on the first answer (an experienced interviewer or oral examiner would always go for the second approach and therefore a CAFA system should do the same).

We can now think of the way in which the student can be asked for input after each experience. The use of assessment methods that can be automatically marked (MCQ, fill in blanks, etc) can generally be termed “convergent” as they ask “closed questions”. This convergence makes it hard to make the reflection influence subsequent experiences - you would need to have a complex hierarchy of questions where the next question depended on the answer to the previous one (as in the example above). Given the difficulty in generating “deep learning” questions of this sort, setting a sufficient quantity to achieve a progression (while giving the students different questions each time) is likely to be impractical for most problem domains.

Therefore a “divergent” approach is needed, where “open questions” are asked and where a student’s answers feed back into later questions. This of course raises the question of how you cope with diverse answers to divergent questions. In the next section we will look at one possible approach.

Simulation

In this section, we will describe the approach we are taking to the problem of generating experiences, gathering the student input and incorporating that input into future experiences.

The approach we have adopted is:

- to use the simulation of virtual worlds as a means of generating experiences;
- to give the student the freedom to manipulate the virtual world (meaning that “what do you want to do next?” is the perpetual “open question”); and
- never to tell the student whether they did what they should have done, but to use the student’s answers to modify the virtual world so that it affects subsequent experiences.

We use the term “Virtual Assessment Environment” (VAE) to describe this arrangement.

Use of computer-based simulation for teaching is well established (Race and McDowell, 1996, §32), but its use for assessment is less well defined in the literature. Brown and Knight list non-computer based simulation as a means of assessment (Brown and Knight, 1994) and Brown extends this list to mention (but not explore) computer-based simulation (Brown, 1999). Race does not include any form of simulation in an extensive list of assessment methods (Race, 2001), while Freeman and Lewis do not mention simulation in a supposedly wide-ranging chapter on using computers for assessment (Freeman and Lewis 1998, ch13). Habeshaw describes some applications of assessment via computer-based simulation, but does not consider the issues involved (Habeshaw et al 1993, §14).

Simulation is well established for assessment in some fields (Habeshaw lists examples in medicine and economics, other include nuclear engineering and pilot training). Here, competence on the simulator is required before you are allowed to train on the real equipment (Cleave-Hogg et al, 2000), so the simulator complements real training rather than replaces it. The simulator facilitates the assessment but does not control it – a human assessor will generate events and will assess the student's response.

There is evidence that use of simulators for learning is not without its problems. Davies argues that the simulator must support the objectives of the learner as well as of the teacher (Davies 2002). Rieber identifies surface and deep strategies that arise when students learn by simulation (Rieber et al, 1996). Pilkington and Parker-Jones point out that students tend to concentrate on the manipulation of simulated objects without obtaining a deep understanding of the principles that underlie the observed behaviour, and argue that choice of task is crucial here (Pilkington and Parker-Jones, 1996). All evaluate simulation as a means of learning, without considering assessment.

Finally, Parush evaluates how learning can be improved if a simulator can record a 'learning history' to allow the student to review their own performance and restart the simulation from any point (Parush et al 2002). This raises the intriguing possibility that a CAA system may be able to assess "process" as well as "product".

First Aid as a case study

In this section, we look at a simulator that has been developed to demonstrate the principles of our approach.

The simulator was developed by one of the authors as an undergraduate final year project for BSc(Hons) "Computer Science and Information Systems". The program simulates the process of cardiopulmonary resuscitation (CPR). The field of First Aid was chosen as it was felt that something subject-neutral would allow the pedagogical issues to be discussed by the widest range of people.

The simulator is not intended to be a "learning tool" helping novices to learn life-saving skills, but is intended to act as a revision aid to people who have

already been trained by other means to check whether they can remember and apply what they learnt. This is analogous to the “computer laboratory sessions” we described in the first section. The academic uses the teaching session to provide support and direction, and the students have access to a CAFA system to allow them (a) to practice the skills covered and (b) to work out if they understand what was covered in the teaching session.

The CPR simulator works in the following way:

1. the simulator displays the condition of the patient and of the lifesaver;
2. the simulator gives the user a list of things they could do (the larger the list the less convergent the simulation);
3. if the user does the right thing then they maintain the condition of the patient, otherwise the patient deteriorates (how soon and by how much depending on the condition of the patient and what the lifesaver has done);
4. as the simulation proceeds, the lifesaver gets increasingly tired and the things they do to the patient have less effect; and
5. the aim of the simulation is to keep the patient alive until an ambulance arrives.

These are intended to be an accurate simulation of the real world, and the developer took advice from the University’s Occupational Health Service and from the British Heart Foundation to ensure the details were medically accurate.

The principles used in its design are as follows:

1. students are free to do anything they want in any order;
2. the student is never told if they have done the right thing, they just see the consequences of their actions (Kolb’s *experiences*) which aims to encourage them to reflect on these when deciding their next action; and
3. the initial condition and responsiveness of the patient can be randomised to ensure that the student gets a slightly different set of experiences every time (therefore different simulator sessions may require different skills)

As an example of a CAA system that takes a convergent MCQ approach, we can compare the simulator with the BBC’s “First Aid Action – Essential Skills – Heart Attack” web site (BBC, 2003). This small site contains four questions with three answers each and claims to provide “a highly engaging scenario” after which “you should have learned some of the most important things to do if someone is suffering a heart attack”.

So far, only informal evaluation has been done (Cafferty 2003), and the approach has been far from rigorous.

Results so far indicate:

- people with no first aid knowledge:
 - get an average of 56% when they take the test the first time;
 - get 100% when they take the test a second time; and
 - get 100% when they take it a third time with the questions covered up.
- people with first aid knowledge get 100% when they take the test the first time.

If these results are confirmed, they indicate that the system is not (in its current form) suitable for formative assessment. Possible reasons are (a) that it does not challenge people who have first aid knowledge but want to verify it; and (b) that people can do well by learning answers rather than by understanding the problem.

Repeating the informal measurements with the simulator (Cafferty 2003), we find that people with no first aid knowledge are unable to maintain the condition of the patient, whereas people with first aid knowledge are able to do so. From observation of the users, it is clear that there are two learning curves involved. The first is learning to use the simulator and, once this is done, the user is able to concentrate on the simulated task and improving their performance.

If these results are confirmed, they indicate that the approach is promising – for this problem domain at least.

The future

So far, we have seen the description of an approach to CAFA that first impressions indicate may be a useful tool.

However, as with other tools, it is only going to be useful in certain circumstances. The role of the academic, as an expert in pedagogy, is to be able to select the right tool for the job and to apply it in the right way. Therefore, the approach must be given a rigorous pedagogical framework, so that its use can be problem-driven (“we’re doing it this way because it’s the right thing to do”) rather than technology-driven (“we’re doing it this way because we can”).

Pedagogical issues that must be addressed include:

- what class of problem is amenable to simulation?
- how can we achieve a divergent interaction between the student and the simulator?
- how can we develop a simulation that is sufficiently realistic while also being implementable, affordable and usable?
- equity of assessment – especially for special-needs students and those with poor IT skills;
- validity and reliability of assessment – are we assessing the student’s ability to do whatever has been simulated, or are we just assessing their ability to use the simulator?

It is also possible that simulation can be used as a tool for summative assessment, for example by renewing someone's "first aider" status based on their performance using a series of simulators. If this is to be done, it must also be justified as "the right tool used in the right way" by the development of a rigorous pedagogical framework. Some extra assessment issues that must be addressed include:

- validity – if the initial state involves randomness then can we be sure we are testing everything we need to test?
- equity – if the initial state involves randomness, are all students getting equivalent tests?
- fairness – by incorporating the student's mistakes into their virtual world, are weak students getting a "double whammy" of getting themselves into a mess and being unable to get themselves out of it?
- plagiarism resistance – how can we be sure who used the simulator? (assuming the assessment is not carried out in an invigilated environment)

Having looked at the rigorous pedagogic framework that is required, we will also need to ensure that it is economically viable, by looking at the "three Ds" of design, development and deployment.

In reality, every simulator is a bespoke piece of software incorporating considerable expertise in the subject domain being assessed. Therefore, design and development according to good software engineering practices will be a large and costly task. We believe that a rigorous understanding of the pedagogy will help, by promoting reuse of software components and focussing software developers on the crucial assessment issues. Nevertheless, a good simulator remains a complex and expensive thing to get right.

There are many aspects to the deployment of such a system that are worth investigation. As well as verifying the validity and reliability of the approach, the attitudes of students and employers needs to be studied (of particular interest is comparing the attitudes of students to a simulator for formative assessment with those to a simulator for summative assessment).

Conclusions

The original ambitious objective that we set for CAFA has been met, at least for our demonstration problem domain:

- *"to allow students to give themselves formative feedback"* – the system allows students to increase their autonomy and to decide the timing and quantity of formative assessment. It provides them with a diagnostic check of how their knowledge and skills relate to a simulated task.
- *"unlimited formative feedback"* – because the formative feedback is not explicitly giving the answer, and because the simulation can have a little randomness, students can use the simulator many times without being able to achieve a good mark simply by memorisation.
- *"without the involvement of an academic"* – the student is autonomous in working out what they do and don't understand, and in the use of reflection, abstraction and experimentation to improve their knowledge and

skills. The academic can therefore devote their time to supporting those students who find their underlying knowledge is insufficient to allow them to do this.

Additionally, through following good assessment practices, it should be possible:

- to assess deep learning from high-order SOLO levels, through the joining together of dependent learning cycles; and
- to provide a high degree of constructive alignment between the learning outcome and the simulated task.

The widespread use of simulation leads us to be confident that the techniques can be applied to a wide range of subject areas. The main potential problem to this is the high up-front development costs that will inevitably mean use is prioritised to those areas where most value is obtained.

We believe this paper shows that simulation has a lot to offer the academic community in providing efficient, scaleable and high quality Computer Assisted Formative Assessment that complements methods such as MCQs.

Acknowledgements

Andrew Young would like to thank Bernard Lisewski, Kath Banks and everyone from the PGCert at Salford University for providing the initial inspiration, and Prof. John Cowan, Prof. John Heywood and the TLDSC for helping to develop the ideas.

Steven Cafferty would like to thank Kate Jones for her medical input to the CPR simulator.

References

BBC (2003) *First Aid Action – Essential Skills – Heart Attack*.
<http://www.bbc.co.uk/health/first_aid_action/es_heart.shtml> (2 June 2003)

Biggs, J. (1999) *Teaching for Quality Learning at University*.
Buckingham:SRHE/OUP

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

Brown, S (1999) *Fit For Purpose Assessment*. Inaugural Lecture, Robert-Gordon University, 21st September 1999

Brown, S. and Knight, P. (1994) *Assessing Learners in Higher Education*.
London:Kogan Page

Cafferty, S. (2003) *Final Year Project – First Aid Tutorial*. University of Salford BSc(Hons) Computer Science and Information Systems, 2003

- Cleave-Hogg, D., Morgan, P. and Guest, C. (2000) *Evaluation of Medical Students' Performance in Anæsthesia using a CAE Med-Link Simulator System*. 4th International CAA Conference, Loughborough, UK, June 2000
- Davies, C.H.J. (2002) *Student engagement with simulations: a case study*. Computers and Education, Volume 39, Issue 3, pp271-282, London:Elsevier
- Davies, P. (2001) *Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible?* 5th International CAA conference, Loughborough, UK, July 2001
- Freeman, R. and Lewis, R. (1998) *Planning and Implementing Assessment*. London:Kogan Page
- Habeshaw, S., Gibbs, G. and Habeshaw, T. (1993) *53 Interesting ways to assess your students*. Bristol:TES
- Kolb, D.A. (1984) *Experiential Learning – Experience as the Source of Learning and Development*. New Jersey:Prentice Hall
- Parush, A., Hamm, H. and Shtub, A. (2002) *Learning histories in simulation-based teaching: the effects on self-learning and transfer*. Computers and Education, Volume 39, Issue 4, pp319-332, London:Elsevier
- Pilkington, R. and Parker-Jones, C. (1996) *Interacting with Computer-Based Simulation: The Role of Dialogue*. Computers and Education, volume 27, no 1, pp1-14, London:Elsevier
- Race, P. (2001) *The Lecturers Toolkit*, 2nd edition, London:Kogan Page
- Race, P. and McDowell, S. (1996) *500 Computing Tips for Teachers and Lecturers*. London:Kogan Page
- Rieber, L., Smith, M., Al-Ghafry, S., Strickland, B., Chu, G. and Spahi, F. (1996) *The Role of Meaning in Interpreting Graphical and Textual Feedback During a Computer-Based Simulation*. Computers and Education, volume 27, no 1, pp45-58, London:Elsevier