Multimedia items in computer-basedtesting

Paper presented at 1994 Australian Association for Research in Education (AARE) Conference, Newcastle, NSW, Nov 27 - Dec 1

John Harvey and David Reid
School of Educational Psychology,
Measurement & Technology
University of Sydney

We should perhaps confess at the outset that this paper was intended to be an account of work in progress, but more accurately, it is still an idea in progress, and our remarks remain speculative.

Test questions have been a key element of computer-based learning (CBL) virtually from its inception (Benjamin, 1988) as instructional designers have capitalised on the machine's capacity to provide automatic scoring, immediate knowledge of results, generate tests from item-banks or from algorithms, and to look after item analyses. Ring (1992, 1993) notes that research indicates that students prefer computer-based testing over conventional testing provided the testing environment and the test itself are both well-designed (1992, p.502).

However, we believe that testing also plays a key role in the acquisition process and the ideas discussed below have application as part of the learning phase as well as in any summative assessment program. That is, we are using the term `testing' in a wide sense.

The idea we are exploring is that it seems to us that although the outward quality of the presentation of CBL software has increased remarkably in recent years (Ambron and Hooper, 1990; Reiber, 1994), and there is a movement towards constructivist perspectives, test items are often substantially `pencil and paper' multiple-choice type items simply transferred to the screen.

There may well be a case that much computer-based testing in fact provides a less flexible environment than orthodox pencil and paper tests in that little provision is made for, say, pre- and reviewing questions, altering answers or deferring answers. Ring (1993, p. 60), on this issue, notes that The American Psychological Association in 1986 issued the guideline that "computerized administration [of tests] normally should provide test takers with the same degree of ... control regarding their responses that they would experience in traditional testing formats" - but, there would seem little evidence of it in the software typically available for schools and colleges.

Little headway also seems to have been made in making adaptive measurement techniques (e.g., latent trait theory) commonly available in computer based testing. Such programs are available (Alessi and Trollip, 1991; Nagaoka, 1992) but generally only at research or other specialist level.

However, these issues are not of immediate concern, although they would need to be if any of the ideas discussed below prove useful and are used in programs which formally assess student performance. Neither is the focus on the development of extended answer (`essay type') questions which require text parsing

algorithms. Rather, at this juncture our concern is to develop and trial test items which make greater use of the facilities available on `multimedia-capable' computers. By multimedia we mean CBL software characterised by the coordinated use of text, graphics, animation, video and sound.

The examples discussed in the next section have come from many sources - existing software, particularly simulation packages, student projects, papers, discussions with colleagues and students. We have chosen these examples as they can be implemented on the computers now becoming common in homes, schools and tertiary institutions.

Example (a)   Making better use of the feedback following a question

Despite what we said above about not dealing with extended text answers, the first example involves such a instance, however, no artificial intelligence is involved. Neither does it require multimedia, but we raise it here nonetheless as an interesting strategy to use with test items in CBL more generally. A question is asked which requires the student to type a short answer of a sentence or two. When this is completed the student clicks a control button which brings up a model answer alongside the student's attempt. The student is then asked to compare his or her answer with the model and to rate how close is the match by assigning a grade. In other words, students are asked to mark their own answers. The format thus incorporates a standard recall test, and a measure of the student's capacity to compare and evaluate. The student's original answer and the rating assigned can be easily written to a file for later use by either the student or the teacher.

If this structure is used as part of the learning cycle in a CBL program, then it lends itself well to exploiting the `expanding rehearsal' hypothesis (Baddeley, 1990). This technique is based on the assumptions that (a), the act of remembering strengthens the response being recalled (provided it is successfully recalled) and (b), the effect tends to strengthen as a function of the the recall interval. It leads to the idea that the interval between re-presentations of a given test item should be a direct function of the strength of the response (and the value assigned by the student to his or her attempt can be used as an index). It works like this\: If an item is answered poorly then it should be re-presented shortly after its initial presentation - and while the student should be able to recall the model answer from short-term memory. If it is then answered correctly the next test of that item is delayed by some amount of time. This cycle continues so that the re-testing interval, ideally, gradually expands - the student is led to recall the answer over longer and longer intervals. If, on re-presentation however, the item is answered incorrectly then the delay is reduced. A discussion of algorithms to implement item queuing is provided in Alessi and Trollip (1991).

Example (b) Usingthe clock

Learned responses sometimes have an important temporal dimension, for example, emergency procedures such as applying artificial respiration, knowing how to set a security alarm, or learning to `speed' read. Computers have a distinct advantage in that the in-built clock can be used to measure a student's response time, or to permit only a given amount of time in which to respond (eg., a student may be given 30 sec to correctly identify the steps to stop severe bleeding), or to control the exposure time of a stimulus item (eg., a picture of a `halt' sign in a program aimed at developing rapid identification of traffic symbols). It can be appropriate with some tasks to use response latency as a measure of strength of learning (Baddeley, 1990), and it can be used as an index for controlling intervals for the item repetition procedure noted in the previous example.

Example (c) Manipulation of objects

The term `direct manipulation' has been used to characterise those interface designs which, amongst other characteristics, use objects (eg., icons) and actions (eg., double clicking) rather than text and command lines (Laurel, 1990). Direct manipulation also can be used in test items. Thus, instead of asking a series of multiple choice questions concerning, say, conductors, capacitors, voltmeters, resistors and the like, a item can require the student to `build' and `test' a circuit by assembling components and using meters from an on-screen `parts box' - in much the same way as graphics can be created by using tools from the palette in a paint program. Apart from the apparent increase in face validity that a direct manipulation item would seem to have, the structure of the item can be so arranged that if a student is, say, unable to get started, then the program can `fill in' that particular part of the answer - that clue perhaps being sufficient to allow the student to move forward and still make good use of the question.

Example (d) Sounds,graphics, video

The previous example suggests how questions might further employ non-text elements such as sounds (eg., a `rough' engine to be diagnosed, or an irregular heart beat, but it could just as easily be a poetry reading, music, film dialogue, or a part of an historical speech), as well as high resolution colour graphics (maps, x-rays, photographs), animations and video clips. For example, a question used in testing office staff on knowledge of emergency procedures might simulate a phoned bomb threat which the person being tested does not expect - it comes among other items on unrelated matters - and is allowed to hear only once. A question then asks the student to recall the content of the call.

Video clips in particular, would seem to have considerable research value for innovative item types as hitherto it generally has not been possible to provide that kind of rich stimulus for a question apart from expensive and/or time-consuming `practical' tests. Thus, a program on power boating could have a short video extract of a boat approaching a buoy in conditions of poor visibility - the clip stops and the student is required to indicate quickly whether to alter course to the left or right. The video then continues and displays the consequences of the choice.

An example of an interesting animation can be illustrated by imagining a schematic of a road intersection which has `cars' converging at different speeds from different distances and directions. The student's task is to estimate the likelihood of a collision. The technique can be used to examine a student's ability to integrate temporal data, an area of cognitive processing overlooked by the types of items found in, say, commonly used IQ tests.

Example (e)Questions in which students generate their own data

It is held that learning will proceed more effectively if learners are actively involved in creating experiences from which they can construct meaning and insight (Shuell, 1992), and teachers go to some lengths to set up learning environments which encourage this. In testing however, usually the questions are completely teacher centred - the student plays no part in constructing any aspect of the item. An interesting departure from this was seen recently in a CBL program on statistics in which the student played a role in generating the data which then was used as the resource for a test item. The program used a simulation of aiming and firing a cannon on an 18th. century warship in which it and the target vessel are rolling in a heavy swell. Although with skill the student could score some `hits', invariably the 20 shots allowed produced a distribution of under- and over-shoots suitable for use in the subsequent question on interpreting data sets. The designer reported that informal evaluations indicated that this seemed to cause increased attention and interest and the users were keen to try the item again.

Example (f) Questions which allow additional resources to be called up

`Open book' questions can be employed by allowing the student to access additional resources via the computer. These could be files which are part of the local program, or links could be made available to servers. As familiarity with the InterNet, for example, becomes more commonplace, and perhaps a critical skill, permitting such access may become a normal part of certain types of examinations in which ability to locate, select and apply information is valued more highly than the ability to recall information. If the files are a part of the program then it is easier to control what resources each student has by way of an "open book" than if they are permitted to bring in their own copies. Similarly, calculators and other instruments can be made available via the software.

Example (g) "Progressive" questions

Problems set in pencil and paper format are generally fixed, in the sense that all the details of a given question are provided at the one time. Question setters work hard to avoid ambiguities and errors. The problem does not build or change over time, yet in `real life', problems are rarely so well defined. With a computer version this can be explored. In a question on a medical procedure, for example, information about the patient can be progressively revealed, the patient can develop additional symptoms, and equipment may fail unexpectedly. Similarly in a question seeking information on how a trainee teacher might cope with a classroom `incident', certain details could be initially withheld and only come to light if the trainee asks to see the student record file (held as part of the program), or after beginning to construct a response, the details of the original trigger incident might be altered, favouring the trainee who did not rush to a hasty judgement. Or, in a test in which a student is required to set up a spreadsheet, a late change in the data might be notified such that students who did not begin to structure their solutions using variables would need to make significant alterations - a forceful way perhaps of driving home

a point.

Ring (1993) has argued that,

the use of constructed response items based on simulation techniques offer students tasks which are more realistic and closer to those they encounter in education and work settings. This work should enhance the face validity of such items; that is, the perception among subject-matter experts and test-takers that such test items are better measures of the test objectives than the corresponding paper-based multiple choice items (p. 63).

There is a need to establish more clearly whether the claims of high face validity of such response types in computerised testing translate into content and predictive validity measures superior to `pencil and paper' questions. This task will be assisted by current research which is developing procedures aimed at getting a better understanding of the cognitive processes used by students engaged on multimedia CBL tasks (Alexander and Frampton, 1994).

References

Alessi S.M, & Trollip, S.R. (1991) Computer-based instruction: Methods and development. 2nd edit., Englewood Cliffs, N.J.: PrenticeHall.

Alexander, M. & Frampton, D. (1994) Technology and thinking: A qualitative study with interactive multimedia products. In C. McBeath & A Atkinson (eds) Proceedings of the 2nd International Interactive Multimedia Symposium, January 4-9, Perth: Promaco Conventions Pty Ltd.

Ambron, S. & Hooper, K. (eds.) (1990) Learning with interactive multimedia. Washington: Microsoft Press.

Baddeley, A. (1990) Human learning: Theory and practice. Hove, UK., : Erlbaum.

Benjamin, L.T. (1988) A history of teaching machines. American Psychologist, September, 703-712.

Laurel, B. (ed.) (1990) The art of human-computer interface design. Addison-Wesley.

Nagaoka, K. (1992). Micro item bank test system: Interactive test construction, test execution and marking-analysis -and-feedback. In ITTE 1992 Proceedings of the 2nd International Conference on Information Technology for Training & Education, University of Queensland, 29 September - 2 October.

Powell, J. The use of multi-media to train intelligent command and control of major fire incidents. In ITTE 1992 Proceedings of the 2nd International Conference on Information Technology for Training & Education, University of Queensland, 29 September - 2 October.

Reiber, L.P. (1994) Computers, graphics, & learning. Madison, Wisconsin: Brown & Benchmark.

Ring, G. (1992) Computer administered testing. In ITTE 1992 Proceedings of the 2nd International Conference on Information Technology for Training & Education, University of Queensland, 29 September - 2 October.

Ring, G. (1993) Developing a computer-administered testing system in a multimedia/hypermedia environment. A paper presented at the Apple University Consortium Conference, Christchurch, August 23-25.

Shuell, T.J. (1992) Designing instructional computing systems for meaningful learning. In M.Jones & P.H.Winne (eds.), Adaptive learning environments. Berlin: Springer-Verlag.