Recent Developments in the Tripartite Interactive Assessment Delivery System (TRIADS)

Don Mackenzie
Professor of Educational Technology in Geosciences
University of Derby

Abstract

The current state of development of the Tripartite Interactive Assessment Delivery System (TRIADS) is reviewed. Recent developments have concentrated on increasing the range and flexibility of question styles so that areas of work that have been difficult to assess by computer previously, such as curved line drawing, can now be tested. Many of the templates currently in development have built-in facilities for randomization of data, images or distracters so that users of the system may take advantage of more open and flexible assessment environments. Such question styles offer the potential to substantially enhance the quality of assessment in areas previously thought to be impossible to test easily on computer.

Introduction

TRIADS is a programmable interactive courseware and assessment toolkit/shell for users of Authorware Professional (Macromedia). The system was initially developed in at the University of Derby as a pure assessment system, based on work that started in 1989 using a DOS-based authoring system (SwanAuthor, Swan Interactive Systems). The assessment system was moved to Authorware Professional in 1992 and has evolved through continuous use with students to its present state of development. During the past three years the system has formed the basis of the Assessment of Learning Outcomes project, funded by the UK Higher Education Funding Council for England (HEFCE) under the Fund for the Development of Teaching and Learning (FDTL) in collaboration with the University of Liverpool (lead site) and the Open University. One of the main aims of the FDTL project has been to evaluate the applicability of computer delivered assessment across the UK Higher Education sector. As a result of this programme, the TRIAD system is currently in use in 32 academic departments and central service units at 19 UK universities in 16 disciplines.

Overview of the System

The TRIAD system is comprised of five principal components:

- Assessment Engine
  The 'heart' of the system, containing all the fundamental procedures required to control the running of an assessment including sign-on, results calculation, results filing and presentation of assessment and module evaluation forms.

- Assessment Configuration Shell
  A series of forms allowing the user to define how the assessment is to run, how the information is to be recorded, how the results are to be displayed and filed and the contents of any menu that may be required.
  An on-screen, run-time configuration facility is also provided.

- Question Sequencer
  An area into which questions are pasted. The sequencer controls question presentation in the appropriate order together with the filing of user responses in an individual file for each question.

- Tutorial Shell
  An area where tutors may place any courseware materials that are coded in Authorware Professional so that they may be accessed as part of the feedback responses to questions.

- Question Templates
  Generic question shells that can be pasted into the question sequencer and filled with question and feedback content by the

tutor. These are provided in a variety of basic screen layouts for each question style. Each template allows finely tuned scoring and offers a variety of modes of operation with full error trapping and facilities for several levels of feedback. Templates for 17 question styles are included in the basic system that is currently under evaluation in the Assessment of Learning Outcomes project and a further 11 are currently at the prototyping stage and on trial in assessments run at Derby. A full list of templates is provided in Table 1

Table 1. Summary of TRIADS question styles and their availability

Question group	Question style	Availability
Point & click	multiple choice	available
	multiple response	available
	Text matrix selection	available
	multiple rectangular hot-spot	available
	multiple polygonal hot-spot	available
	multiple clickable object	available
	multiple true/false	available
	assertion-reason	available
	plot point - check X-Y	available
	line/word clicked	under development
	multiple question - multiple answer	under development
	matching pairs text/graphic	prototype
Move object	label diagram	available
	random sequential label diagram	available
	random image label diagram	prototype
	build diagram	available
	classification	available
	simple sequencing	available
	multiple sequence classification	prototype
	multiple-sliders on scales	prototype
Text-numeric entry	single & multiple text entry	available
	single & multiple numeric entry	available
	mixed text & numeric entry	available
	fill a table - text/numeric	under development
Draw object	draw straight line	prototype
	draw arrow	prototype
	draw box	under development
	draw circle	under development
	draw curve	prototype
	multiple versions of above	under development
Combinations	matrix selection -sequencing	available
	matrix selection - classification	prototype
	multiple text entry - labelling	prototype
	multiple text entry - sliders	under development
	multiple choice - matching pairs	prototype
	plot graph - draw line - enter values	prototype

Complex questions using combinations of most of the question styles listed are possible. Tutor configurable Module and Assessment Evaluation Forms are also built into the system.

Popular misconceptions about computer-based assessment

Whilst many would concede that computer-based assessment has a significant role to play in a formative environment, the method is often dismissed for summative purposes because it is perceived that:

1. CBA is able to deliver multiple choice or similar simple styles of questions only;
2. the results are unreliable because these simple question styles are susceptible to scoring by guesswork;
3. CBA is applicable only to low level testing of factual knowledge;
4. CBA is hardware resource intensive;
5. there are serious problems associated with validation of user identities and security of results.

The recent developments of the TRIAD system that have concentrated on tackling some of these issues are outlined in the sections that follow.

Increasing the range and flexibility of question styles

One of the advantages of using software authoring systems to produce assessments lies in the wide range of interactivity that they provide, coupled with the ability to mix different styles of interaction on a single screen. This has allowed us to achieve our original aim of moving away from the multiple-choice style of questioning with its inherent problems. It is the intention of the TRIADS team that, ultimately, the system will provide the tools to test most aspects of most disciplines given the presence of question designers with sufficient imagination.

Reducing the guess factor - Move Object question styles

TRIADS templates for 'move object' question styles have a high degree of flexibility built into them. The simplest form of move object question might relate to the labelling of a diagram. This question style, if badly constructed (e.g. four labels and four parts to label) has the potential to give a high score purely by guesswork on the part of the student. The TRIADS label diagram templates however allow the guesswork score to be minimised by providing the tutor with the facility to:

- include labels for which there are no positions;
- include positions for which there are no labels;
- include positions that can score more than one label;
- score individual labels in more than one position;
- negatively score particular labels in particular positions
- negatively score any label in an inappropriate position.

Up to 26 text or graphic labels may be on screen at any time and the ability to 'fine tune' the scoring means that a wide spectrum of scores between 0 and 100% will normally be achieved by a cohort of students.

Other varieties of move object question styles allow the tutor to test more than just knowledge. Varieties of sequencing questions are particularly good for this. In many areas, aspects of understanding revolve around sequences of processes or actions, or placing arguments or evidence into sequence to demonstrate a point. As with simple labelling types of question, dummy items may be included. Tutors may select from four scoring algorithms in TRIADS and a further algorithm that allows for a range of correct sequences to be tested is currently in development. Analysis of student's responses to carefully designed questions of this type can give a substantial insight into areas of understanding and misunderstanding in a student cohort that can be used to pinpoint topics for targeted tutorial work.

Getting more out of selection - Point and Click question styles

Whilst multiple choice questions may have a degree of usefulness in formative assessment, they possess a number of serious disadvantages when used summatively. Even if the difficulty of writing good quality distracters is overcome, the high potential for the student to guess the correct answer means that the resulting score is of limited usefulness to the tutor. Even if the student scores 100% on a multiple choice question, the tutor cannot tell whether the student had thought deeply about the answer and worked it out from first principles, knows the answer or has luckily guessed the answer. In such assessments where 1 selection out of 5 is the norm, such problems cast substantial doubt upon the validity of the results and pass marks need to be increased in order to compensate for these shortcomings.

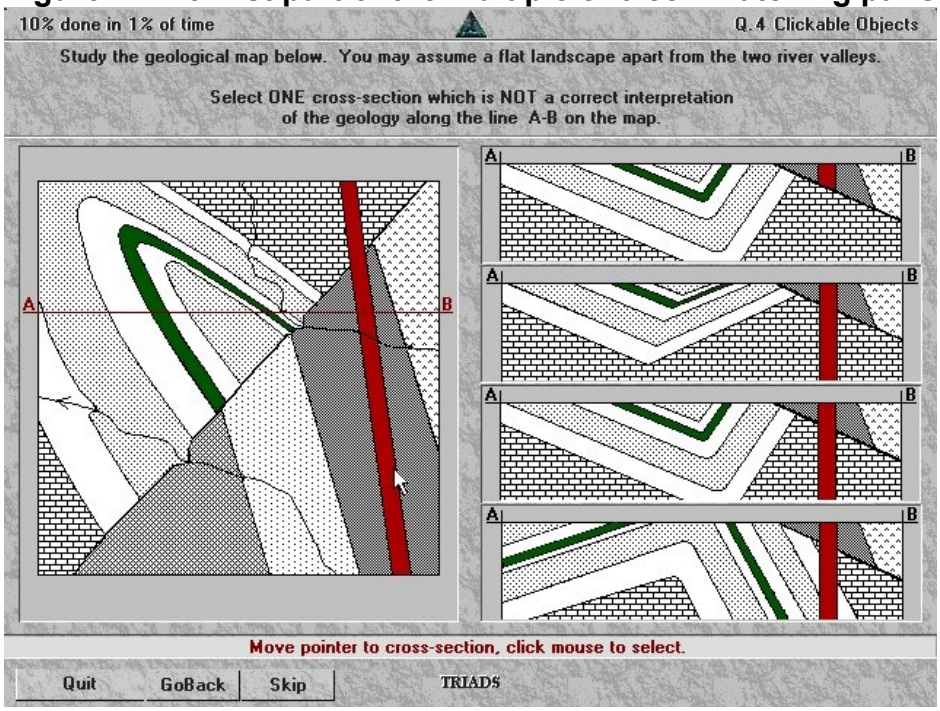
This situation may be improved by the use of multiple response question styles (where more than one selection in an array is correct) with an increased number of distracters. The guess factor can be substantially reduced by negatively scoring incorrect selections and leaving the question open-ended, i.e. giving no indication of how many selections are correct. Any overall negative scores resulting from such questions can be set to zero on exit from the question.

The existing TRIAD system provides a single set of templates for these question styles that can be configured to constrained or open ended multiple response types or to simple multiple choice styles for text and/or graphics.

One of the principal disadvantages of questions of this type when used in their traditional manner is that they give little feedback to the tutor concerning the level of knowledge or understanding of the student. However it is possible to get more information by adding supplementary questions to the initial multiple choice. An example of such a question style is illustrated in **Figure 1**. This question is a modification of a simple multiple-choice question style in which the student was asked to identify the correct

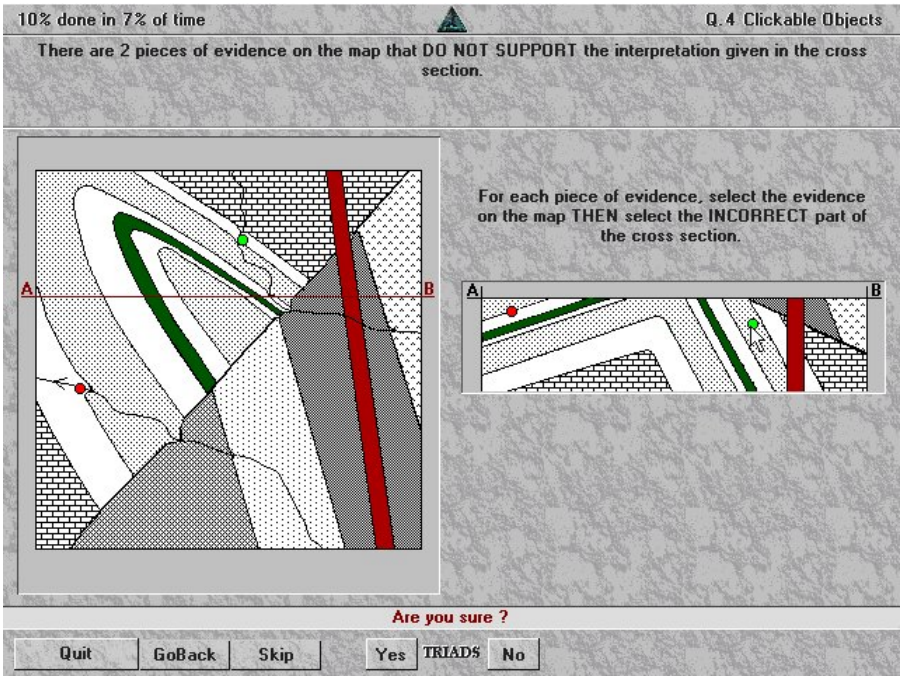
geological cross section from an array of cross sections. (courtesy Open University).

Figure 1. The first part of the multiple choice - matching pairs question



In its original version there were 8 cross sections as distracters which substantially reduced the guess factor. However it is possible to improve on this question and gain more information about the student's understanding of geological maps by altering it slightly. In the modified version shown in **Figure 1**, the student is asked to select any one of the cross-sections that is *incorrect*. Upon selection of a cross section they are presented with a 'matching pairs' style of question that asks them to identify the features on the map that do not match with the cross section and thus provide evidence for the cross section not being correct. (**Figure 2**).

Figure 2. The second part of the multiple choice - matching pairs question



There may be more than one such feature for the cross-section selected. On completion of this part of the question, the screen returns back to the original multiple choice screen which now displays the remaining cross-sections. The selections and matching pairs are continued until the correct cross-section remains, whereupon the question finishes. The question automatically finishes if the correct cross section is selected prematurely.

This question style overcomes many of the disadvantages of traditional Multiple choice styles and mimics the way that any student who was seriously trying to identify the correct section would work. The fact that there are only four items at maximum on screen from which to select, together with the fact that the question finishes if the student selects the correct cross-section prematurely, means that there is a high chance that any student trying to guess the answer will be ejected from the question at an early stage. The TRIADS template for this style of question will also allow as many cross-sections as defined by the tutor to be selected at random from a bank of cross-sections. The selections displayed will always contain the correct cross section.

In the example given above, if each cross-section is related to a particular learning outcome, it is possible for a tutor to evaluate the scores for the cohort of students for each learning outcome. Analysis of the scores for each learning outcome may allow follow-up tutorials or modifications to course content to be more focused.

This template could modified by replacing the graphics with text and used, for example, to test the student's ability to evaluate the validity of argument in a piece of prose (equivalent to the map in the example given) against statements relating to the prose (equivalent to the cross-sections).

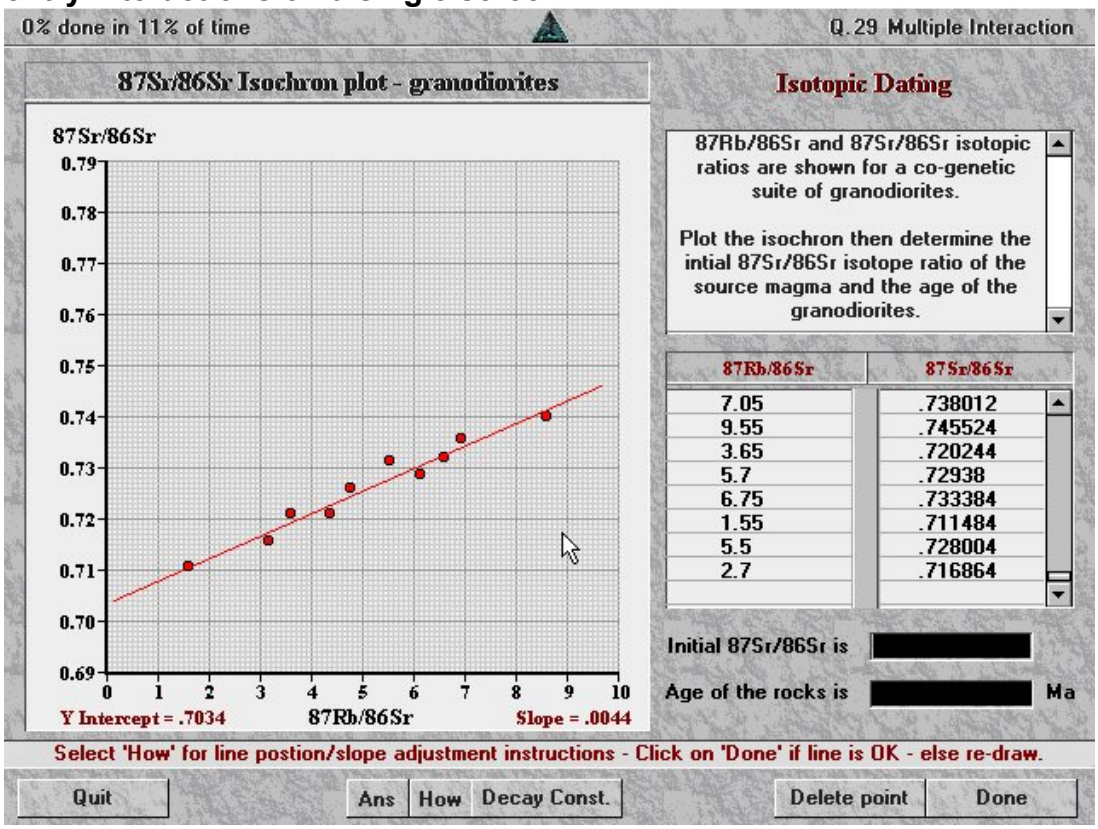
Complex question styles and simulations

The TRIADS system allows the use of complex questions up to the level of full simulations. Anything that can be coded in Authorware Professional or even the results of operations in an external program can be evaluated as part of an assessment in TRIADS. The use of multiple interaction styles within a single question is one of the great advantages that a software authoring system provides. Many assessment systems often only give limited use of a single interaction style at a time.

An example of a complex question style is illustrated in **Figure 3**. The question requires the student to plot the data points on the graph and draw a best fit line by eye through the points. By measuring the slope and intercept of the line and applying a simple formula, the student should be able to estimate the initial strontium isotope ratio of the melt that gave rise to the rock samples analysed and to calculate the age of the rocks. The numeric answers are entered at the bottom right of the screen. This question style continues the theme of randomization since all the data is randomly generated along a line with a slope and intercept that is randomly generated within sensible limits and has a random scatter about that line. Each time the question runs, the student is provided with a different dataset to plot so that the calculated isotope ratio and age will be different. The template checks that the student has plotted each point in its correct position on the graph and checks the position and slope of the line drawn and allocates a score accordingly. The numerical entries are assessed against an answer calculated within the template for the data generated by the randomizer. The tutor may vary the error limits within which a correct score is achieved and also the proportion of the marks allotted to plotting and to calculation. This template is still in the prototype stage but when complete will allow the tutor to define different types of graph paper including linear, log-log, log-linear, probability and triangular in order to provide the widest possible range of applicability.

A variant of this question style will allow the student to enter their own data so that the results of laboratory and field exercises may be automatically assessed.

Figure 3. A complex question style with two graphic and two numeric entry interactions on a single screen.



Question styles that other systems cannot reach!

A longstanding problem in computer-delivered assessment has been that of assessing free drawn curved lines. In science it can be useful to test the position and shape of such curves. This semester saw the first trial run of such a question style with a group of Stage 1 Earth Science undergraduates.

The question screen is illustrated in **Figure 4**. Students are asked to use the information provided to estimate the position on the topographic map of the surface outcrop of a coal seam. Given that this is the intersection of an inclined plane with an irregular surface, the answer will be a curved line. Part of this line has been completed on **Figure 4** as an illustration. When setting up the question, the tutor may divide the line into a number of segments and allot a score to each segment, the sum of scores for all segments totalling to 100 for the question. In this way, more critical parts of the line can be scored more highly. Erroneous segments may also be defined and negatively scored to trap common errors of misconception by the student. Feedback to the results file (and to the student if activated) gives a score for each segment so that missed segments can easily be seen. The anonymous results for one of the students answering this question in the recent run is shown in **Table 2**. Additional graphic feedback of the correct answer may also be provided on screen to the student. Work is continuing on this template to allow the plotting of multiple lines in different colours.

Figure 4. Example of a ‘draw curved line’ question style

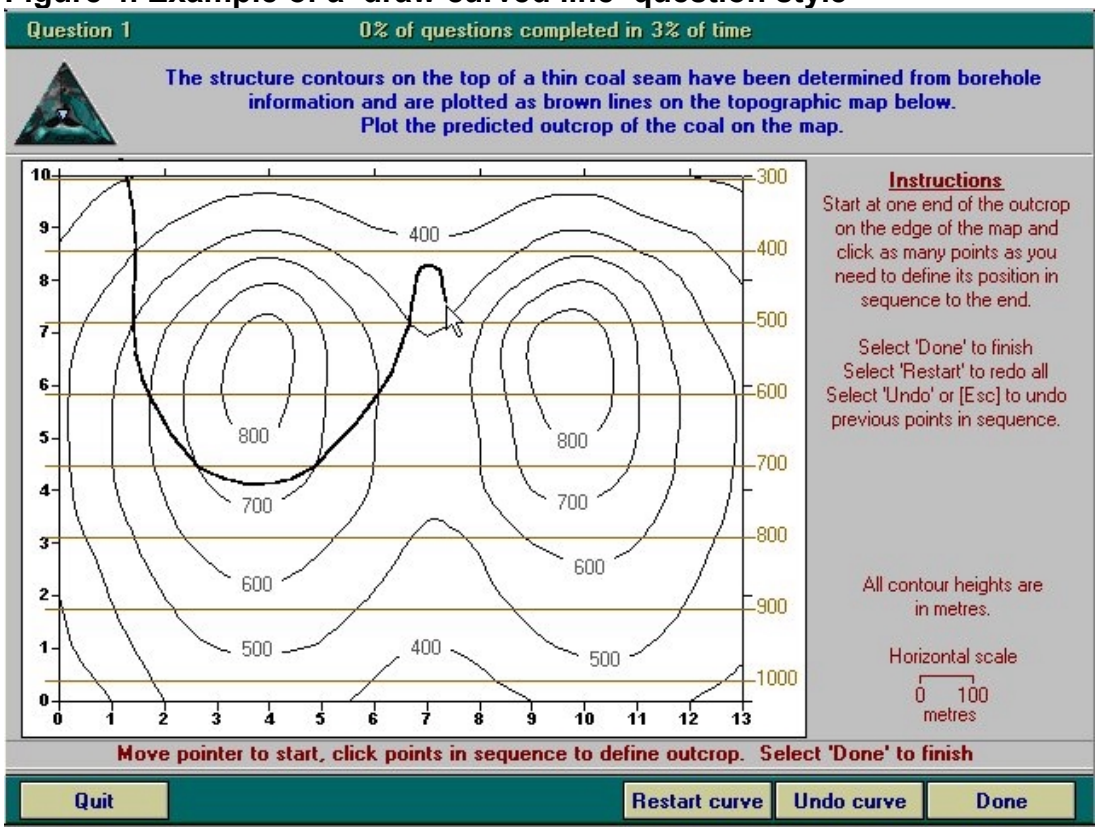


Table 2 Question results from the Draw Curved Line question style

TRIADS - Geological Field & Map Skills - Year 1 - 2nd Assessment 1999

Results for: JOHN THE ANONYMOUS born 1/3/79
FINAL SCORE FOR THIS ASSESSMENT IS 60 %
Assessment started at : 15:32:21 on 14/05/99
Time taken for assessment : 11.45 minutes.

DETAILS	%	time	wt	ac%	
Draw the outcrop		70	126	1	4

Question
The structure contours on the top of a thin coal seam have been determined from borehole information and are plotted as brown lines on the topographic map below.
Plot the predicted outcrop of the coal on the map.
Correct answer(s)
300w
400w
500w
600w
700w
700spurw
700cw
600cw
500cw

valley
500ce
600ce
700ce
700spure
700e
600e
500e
400e
300e

Answer(s) given
300w = 100% x .05
400w = 100% x .05
500w = 100% x .05
600w = 100% x .05
700w = 100% x .05
700spurw = 0% x .05
700cw = 100% x .05
600cw = 100% x .05
500cw = 100% x .05
valley = 0% x .1
500ce = 100% x .05
600ce = 100% x .05
700ce = 100% x .05
700spure = 0% x .05
700e = 100% x .05
600e = 100% x .05
500e = 100% x .05
400e = 0% x .05
300e = 0% x .05

It is interesting to note that on the first run through with students, this question produced a very good range of scores, with a mean of 51% and a correlation with the final scores for the whole assessment of 0.68, the highest in the test. Clearly a highly discriminating question style. When completed and fully tested this style will open the way to many new areas of assessment.

Question styles for the arts and humanities

Development of the TRIAD system to date has concentrated on the styles of questions that are typically required by science courses and the analysis of text has taken a back seat. This situation will be rectified shortly with the advent of a number of new question styles that are currently on the TRIADS 'drawing board.' Although it will not be possible for some time to analyse free text entry of long answers with any degree of confidence, there are other ways of analysing text. For example it is possible to ask the user to identify and/or edit sections of a larger piece of text in answer to questions provided by the tutor. In languages, the selection of bad grammar,

inappropriate use of words or tense can be coupled with a text entry editing procedure that allows the student to enter the corrected phrases. Copy/paste facilities are already available in the latest TRIADS multiple text entry templates. These templates will also allow the student to copy text from external sources to paste as answers into TRIADS questions. All TRIADS templates are easily modified to take sound files and video-clips.

Extending the flexibility of computer-based assessment – randomisation

The hardware intensive nature of computer-delivered assessment may preclude its use for summative assessment in many institutions where rooms equipped with large numbers of machines are not available to conduct formal tests or examinations in the traditional sense. However there is a middle way that may be of more widespread applicability if we carefully examine our motives for assessing students.

So why do we assess students?
A number of principal reasons may be suggested:

- to aid student learning – formative assessment;
- to grade and rank student performance - summative assessment;
- to monitor student progress - scored formative assessment;
- to monitor the effectiveness of teaching and learning strategies;
- if it isn't assessed, the students won't do it!

In recent years there has been a move away from assessment by final examination alone. Continuous assessment strategies are now widespread and the advent of the semesterized modular schemes into UK universities means that students are examined on smaller areas of work, more frequently than in the past. Students now come from a wider range of backgrounds and personal circumstances than ever before. Many are working while studying and the number of mature students with families is also increasing. These changes mean that fixed time tabling associated with such courses can create problems of accessibility. Computer-based assessment can help to ameliorate some of these problems by providing assessments in a more open environment than hitherto.

Formative assessment as a learning tool

Clearly the provision of formative assessments as an aid to learning now creates few difficulties and can be very effective, particularly in combination with courseware provision. An example of a question style that takes advantage of full feedback provision is illustrated in **Figure 5**. This question is designed to test the student's ability to relate real features to textbook diagrams. So often, textbook diagrams are stylised representations of the real materials that bear little resemblance to the real thing. In this question, images of volcanic features selected at random from a bank of images are presented one at a time to the student who is required to identify the feature and move the appropriate label to the correct position on the diagram.

When run in tutorial mode with feedback activated, the template will automatically generate a summary feedback screen on how the user has performed (**Figure 6**) and present a 'View Answers' button that allows the user to scan through the images accompanied by descriptive text entered by the tutor. (**Figure 7**). This level of feedback makes this question style a valuable learning tool. This style of question could be applied in many disciplines and is especially useful in biological applications.

Figure 5. A random image label diagram question style

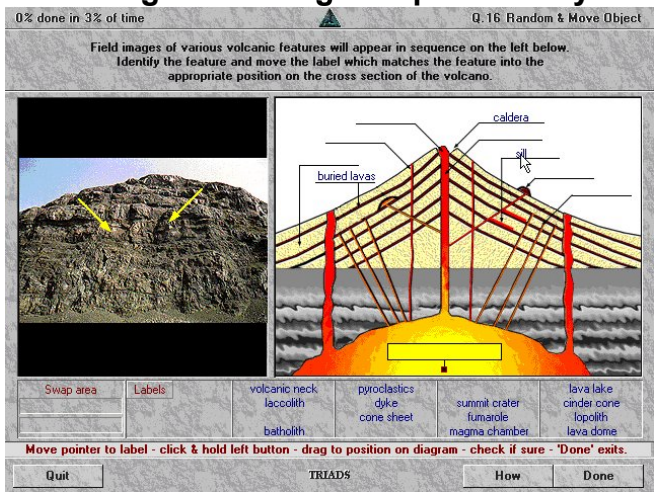


Figure 6. Automatic feedback generated by the random image label diagram question style

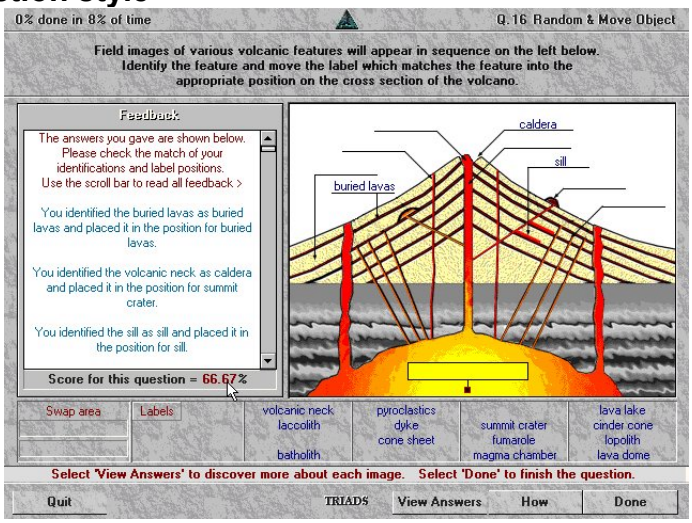
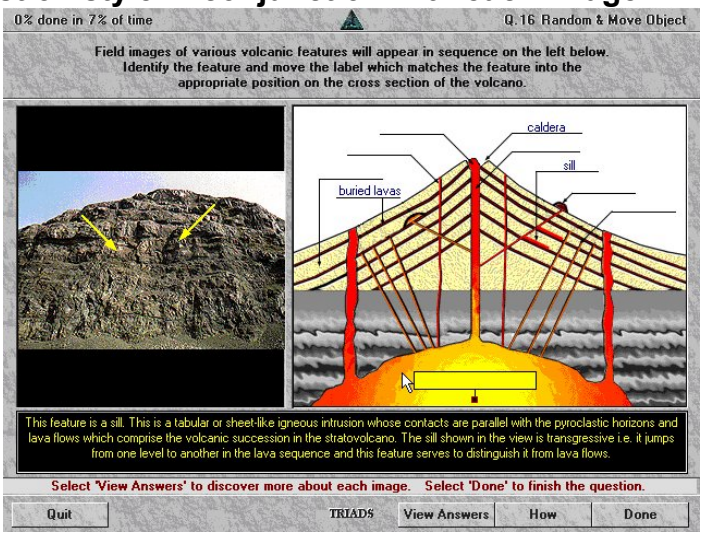


Figure 7. Tutor defined feedback displayed by the random image label diagram question style in conjunction with each image.



In addition to feedback built into individual question templates, the TRIADS system contains a Tutorial Shell section to enable the provision of very detailed feedback up to the level of full courseware in response to a student's answer to a question. Future versions of the engine will allow the system to be configured to run either in Assessment Mode or in Tutorial Mode. In the former, the system will run the assessment and jump to the appropriate tutorial sections as required if a low score is achieved on a question. In Tutorial Mode, the system will run the tutorial, via a menu if required, and allow the facility to pull in questions from the assessment section at any point in the tutorial. The assessment engine will keep track of the scores for all the questions asked in the tutorial. When fully developed, this will represent a complete courseware/assessment shell solution for courseware developers.

Scored Formative Assessments and randomization

The provision of computer-delivered summative assessments in a formal examination situation can create difficulties of hardware availability. But, do we really need to examine students in this way? Is it possible that many summative assessments could be replaced by scored formative or open book, 'staged assessments' throughout the semester? Many of the new TRIADS templates will have the facility for the provision of randomized data, images or distracters selected randomly from a bank such as those already described. The system itself also has the capability to allow the random selection of questions from groups of questions and will eventually allow the random selection of questions from a bank of questions for the whole test. The latter method of randomization should however be used with care since it is difficult to assemble a large number of questions with similar degrees of difficulty and maintain the same balance of topic coverage for each student when questions are selected from a single large bank.

The use of randomization allows 'time windowed', open book style assessments to be provided anywhere on campus. Each time the assessment is run, it will contain questions about the same topics but with different data, images or distracters. Even if students complete the assessments in small groups, they will all have learned a great deal by the time they had done several since few of the questions will be the same. That, of course, is one of the purposes of assessment.

Such assessments can be classed as ‘scored formative’ type and cannot count as fully summative. Although not a complete solution to ranking of students, it could go a long way to relieving the pressure on academic staff of the marking associated with more traditional forms of coursework assignments. Scored formative assessments run in this way have the advantage that once all students have completed the test, a simple run-time configuration will allow the assessment to be viewed again by all students with the feedback activated.

Clearly there may be a problem associated with validation of user identities associated with this type of assessment but it is no more of a problem than with word processed essays that are a common form of coursework assignment at present. Users taking computer-delivered assessments that are run under exam conditions can be validated in the same way as for a paper-based examination.

Conclusions

It must be confessed that one of the prime motives for many who use computer-delivered assessment is the notion of the time saving that is potentially on offer. In order to achieve such time savings it is tempting to populate assessments with large numbers of hastily created multiple choice questions that give no more than a semi-quantitative estimate of student performance and a very low level of information back to the tutor.

To concentrate on the time saving aspect of computer-delivered assessment is to ignore the many other advantages that this method of assessment can offer. By using some of the more complex question styles such as those described above, it is now possible to deliver both broad-based and in-depth assessments of areas that were previously thought to be impossible to test on computer. Many of these more complex question styles can give very detailed feedback concerning how the question was attempted and on areas of misconception held by the student. This allows the tutor to modify courses or to target tutorial sessions more accurately and give a real boost to the quality of learning and teaching.

The objectivity of computer-delivered assessments means that the results obtained from them are much more suitable for statistical analysis than those from most traditional assessments. By maintaining assessments constant

over a period of time, it is possible to use them monitor the effects of changing modes of course delivery and other factors that may affect student performance.

Randomization of questions and question content means that computer-based assessments may be delivered in a much more flexible and open manner. This meets the needs of today's students better than some traditional methods. Improvement in technology now allows assessments to be delivered anywhere on or off campus.

Most importantly, well designed computer-delivered assessments have the potential to enhance the quality and scope of assessment in many areas. It is however more difficult to design a good computer-based assessment than it is to set an essay title. The payoff comes in questions and assessments that have a long shelf life and the potential to save staff time in the longer term.

It is of course true that not all areas can yet be tested by computer, but the TRIADS team are working on it!

Acknowledgements

The help that the TRIADS team, Professor Chris Paul and Dr. Alan Boyle at the University of Liverpool, Dr. Dee Edwards and Dr. Dave Williams at the Open University and Dr. Dave O'Hare have given in their support, educational advice and promotion of the system across the UK is gratefully acknowledged. Particular thanks go to Helen Wilkins who has undertaken much of the technical maintenance and upgrading tasks that are necessary as templates develop through use with students.