A computer-aided continuous assessment system

B. C. H. Turton

School of Engineering, University of Wales Cardiff

A high-quality assessment system should have the following attributes: rapid feedback to the students, appropriate and detailed feedback, and an effective grading system which provides an accurate overall grade as well as information which identifies the student's weak areas. As staff/student ratios worsen, providing such a system will become more difficult and consequently computer assistance in this task is becoming more attractive. This paper describes a Computer-Aided Assessment (CAA) system based on a modified version of the multiple-choice questionnaire. The CAA has been designed to be used in continuous assessment, with features that discourage plagiarism and provide appropriate feedback. Over a hundred students were tested using this CAA and the results were compared with a more traditional assessment system. In addition, questionnaires were used to assess the student's reaction to the CAA. The results were highly satisfactory, and a more advanced version of the original software is under consideration.

Introduction

Universities within the United Kingdom have had to cope with a massive expansion in undergraduate student numbers over the last five years (Committee of Scottish University Principals, 1993; CVCP Briefing Note, 1994). In addition, there has been a move towards modularization and a closer monitoring of a student's progress throughout the year. Since the price/performance ratio of computer systems has continued to improve, Computer-Assisted Learning (CAL) has become an attractive option. (Fry, 1990; Benford et al, 1994; Laurillard et al, 1994). To this end, the Universities Funding Council (UFC) has funded the Teaching and Learning Technology Programme (TLTP). However universities also have a duty to assess as well as to teach. This paper describes a Computer-Aided Assessment (CAA) system capable of assisting in grading students and providing feedback. In this particular case, a continuously assessed course (Low-Level Languages) of over 100 students is considered. Typically, three man-days are required to mark one assessed piece of coursework from the students in this class. Any feedback on how the questions were dealt with by the student are of necessity brief. Most of the feedback is provided in a tutorial session that covers the pitfalls encountered by the majority of the students.

A CAA solution was sought which covered the following points:

- rapid feedback;

- appropriate feedback;

- breakdown of the strengths and weaknesses of the student;

- breakdown of the strengths and weaknesses of the class

- a system that makes plagiarism difficult;

- effective assessment of the student;

- reasonable resource implications.

Rapid and appropriate feedback is clearly necessary for a course lasting only 12 weeks. Any assessment method will obviously be more useful for teaching purposes if it can form a profile of the student and the class. Plagiarism must be guarded against since continuous assessment precludes examination conditions. All these conditions must be met within the resources available.

The Computer-Aided Assessment system

In order to facilitate rapid marking, a fixed set of solutions was deemed to be the only cheap and effective means. Traditionally, this has been thought to be a rather limited system as feedback can be poor. In order to combat this problem, each available choice within a question was designed to capture a particular misunderstanding of the subject. This discipline was highly informative as it requires the examiner to consider the common errors made by students. By storing information on common mistakes, a detailed explanation can be given to the student which deals with the specific area of misunderstanding that led him or her to choose a particular, inappropriate answer. An explanation of the correct answer is also given to ensure that those who have guessed the answer can still benefit from the correct explanation.
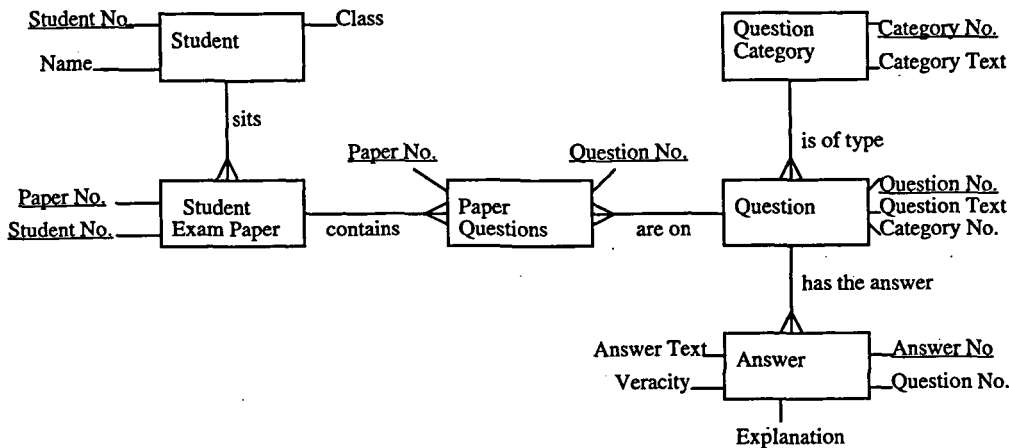

Figure 1: Entity Relationship diagram for CAA

Each question must also be classified so that a summary can be made of the student, and classes, for different categories. Such classification enables the computer to create summaries for students and classes. Figure 1 shows an entity/relationship diagram for the situation as described so far.

The key disadvantage of this system is its susceptibility to plagiarism. Unlike the case with traditional assessments, plagiarism would be impossible to detect between students taking the same paper. An alternative would be to set different questions for each student, but it would be impractical to design a paper with 10 different questions for over 100 students. One solution would be a bank of multiple-choice questions from which 10 questions are selected for each student; however, this bank would have to be very large, of the order of 100 questions, to ensure that on average any two papers would have no more than one question in common. In addition it would be difficult to ensure that a large range of questions maintained roughly the same difficulty and subject coverage.

The solution adopted was to have a small bank of questions from which 10 were selected for each student. Each question has a number of different ways of being asked, a number of correct answers, and a number of false answers. The question asked of the students is one instance of the text for that question with one true answer, a number of randomly selected false answers from the set of false answers to that question, and a 'don't know' answer. This combines the advantages of having a small set of questions yet ensures a variety of combinations such that plagiarism is discouraged. For example, a simple question designed to check a student's ability to perform base conversion could be stored as follows:

Choice

Question version 1: Which value is equivalent to the number 112_5.

Question version 2: $X-1012_3 = 0$, select X from the choices below.

Answer (true) 100000_2

Answer (true) 44_7

Answer (false) 112_{10} or 1012_{10} (student failed to perform conversion, guesswork)

Answer (false) 332_3 (student failed to notice digit range 0–2 for base 3)

Answer (false) 25_6 (student transposed the digits)

Question

There are twelve combinations for a simple three-option multiple-choice questionnaire; one possibility is:

$X-1012_3 = 0$, select X from the choices below.

(a) 25_6

(b) 44_7

(c) 332_3

Assume q questions are required for m students, with c choices per question not including 'don't know'. A formula will be given for each case followed by a numeric result assuming 100 students, 10 questions and four choices.

Scenario 1: Entirely different questions for each student

Probability of two papers not duplicating any question	= 1.00
Number of questions to be prepared = $q.m$	Example = 1000 questions
Number of answers to be prepared = $q.m.c$	Example = 10000 answers

Scenario 2: A bank of questions (b questions in the bank, assume b=100 for the example) Probability of two papers sharing s questions, = $_{b-q}C_{q-s}\,_qC_s$ / $_bC_q$. Figure 2 shows the results for the example numbers.

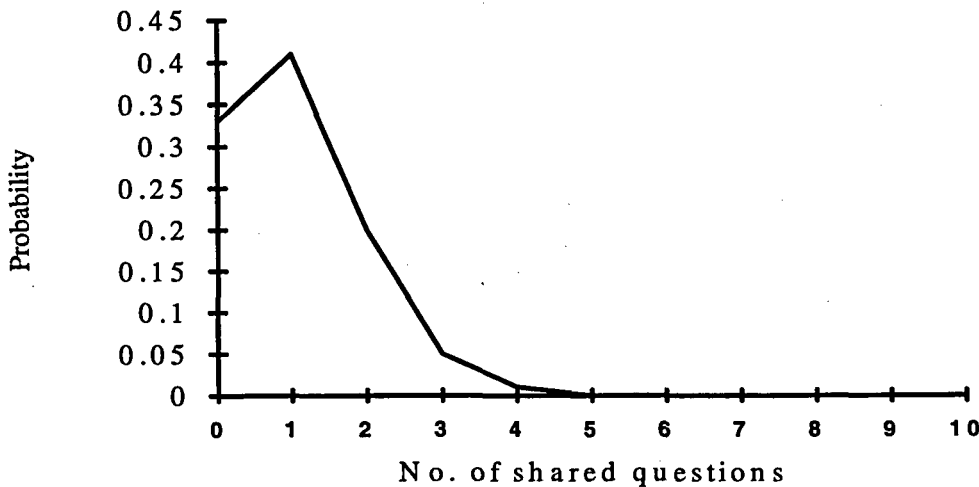|	
Number of questions to be prepared = b	Example = 100 questions
Number of answers to be prepared = $b.c$	Example = 400 answers
	(4 per question)

Figure 2: Probability that two papers will share zero or more questions

Scenario 3: In this scenario, a distinction will be drawn between the whole question (questions and choices) and the text of the question itself (not including choices). The whole question will be referred to as the 'question', and the text of the question itself will be referred to as the 'question-text'. In this scenario, there is a bank of b questions with n versions of the question-text for each question. There are t true and f false versions of answers to that question from which c choices are given to the student, only one of which is true. A variety of question-texts and answers can be used to create a single question with c choices of answer. Clearly some combinations of answers and question-texts will be so similar that the questions are for practical purposes identical, while other combinations will be very different. In order to provide a fair comparison with the other techniques, questions which are very similar should be considered to be the same question. Two methods of grouping questions together as being essentially identical can be defined:

Option (i): If two questions share the same 'true answer' choice and the same question-text, they will be deemed to be essentially the same question.

Option (ii): If two questions share the same question-text and choices, in any order, they will be deemed to be essentially the same question.

The first option assumes that a student is cheating by acquiring someone else's 'perfect' paper in addition to his or her own, and so can copy where the question-text and the true answer appear on both papers. The second option is strictly correct in that the questions are different if the choices or question-text are different. The author of this paper prefers option (i) as a measure of similarity, as it more closely reflects the possibility of plagiarizing another student's work. In the equation that follows, the number of distinct versions of question is denoted by the symbol v. This number can be calculated as follows: for option (i) $v=t.n$; for option (ii) $v=t . _{(c-1)}C_f . n$ (see Figure 3). The value of v will be substituted in the probability formula given later.
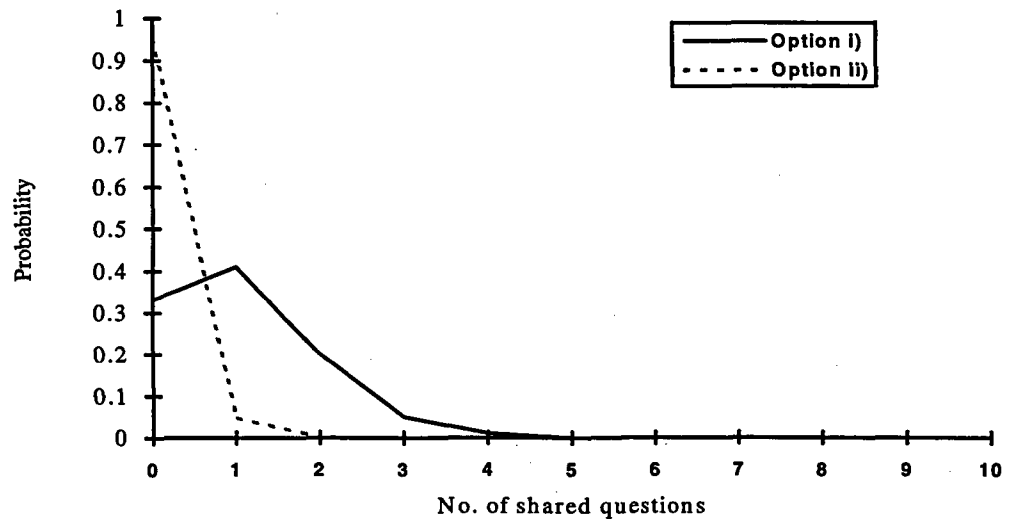

Figure 3: Probability of shared questions {b=25, q=10, v=4, option i) v=4, option ii) v=80}

The formula for calculating the probability of two papers having i questions in common, of which s are the same variant and consequently can be easily copied, requires some explanation. Figure 4 represents the simple case where there is only one version of each question.

In this case, the formula is simply $(_{b-q}C_{q-s} . {_q}C_s)/{_b}C_q$. The reader must remember that there is a bank of b questions of which q are on the paper which is being used for plagiarism purposes. So $_{b-q}C_{q-s}$ represents the number of ways of choosing $(q-s)$ non-duplicated questions from the possible $(b-q)$ non-duplicated questions in the question bank. $_qC_s$ represents the number of ways of choosing s-duplicate questions from the set of q questions which could be duplicated. So the total number of ways of producing a paper which has exactly s duplicated questions is $(_{b-q}C_{q-s} . {_q}C_s)$. Since the total number of ways

Bank of b questions

q questions on the first paper which the student is attempting to copy from

Paper 1

s same questions

Paper 2

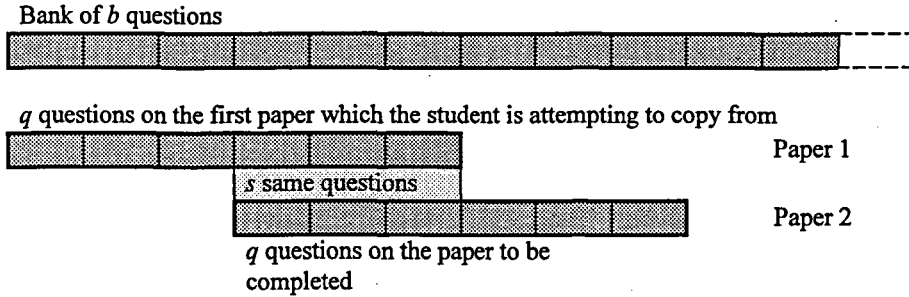q questions on the paper to be
completed

Figure 4: Pictorial representation of questions chosen from a bank

of producing q questions from b is $_bC_q$, the probability of doing so by chance must be $(_{b-q}C_{q-s} \cdot {}_qC_s)/_bC_q$. In scenario 3 this is complicated by the fact that even if a question is duplicated, it may be a different variant to the one chosen on the original paper. Consequently, it will be assumed that of the q questions, i are the same question and s of them are the same variant of the same question.

Bank of b questions

q questions on the first paper which the student is attempting to copy from

Paper 1

i shared questions

s-shared
questions
same
version

Paper 2
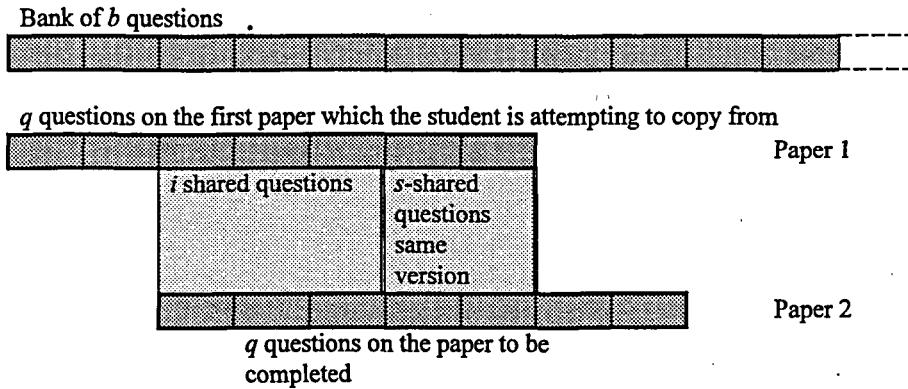
q questions on the paper to be
completed

Figure 5: Pictorial representation of questions chosen from a bank with different versions of question

The formula must now be split into three sections: the number of ways of choosing $(q-i)$ non-duplicated questions from $(b-q)$ questions, multiplied by the number of ways of selecting $(i-s)$ duplicate questions but different variants from q questions, times the probability of that event, multiplied by the number of ways of selecting s questions from the remaining $(q-(i-s))$ questions, times the probability of that event. Now the equation must be summed for all possible values of i, and divided by the total number of ways of choosing q questions from b. This gives a probability of any two papers sharing s 'identical' questions of:

$$P(b,v,s) = \sum_{i=s}^{q} \left(\frac{_{b-q}C_{(q-i)} \cdot {}_qC_{(i-s)}\left(\left(v-1\right)/v\right)^{(i-s)} \cdot {}_{(q-(i-s))}C_s \cdot \left(1/v\right)^s}{_bC_q} \right)$$

Assume for this example that there are two versions ($n=2$) of the question-text for each question, two true answers ($t=2$), and six false ones ($f=6$) from which four options ($c=4$) are chosen. Finally, let the bank have 25 questions ($b=25$). A particular question may only appear once on a particular paper, irrespective of the way it is expressed.

Number of questions to be prepared $= b.q$ Example = 50 questions
(25 questions each with two different question-texts)

Number of answers to be prepared $= b.(t+f)$ Example = 200 answers
(Eight for any one question, six false, two true)

In order to ensure that the frequency of a particular choice being on a question is the same for both true and false answers, $t \approx f / (c-1)$ which is true for the example given.
Once this bank of questions and answers has been created, there is no requirement to create a new bank for the next year, as comparison of papers between students from different years will not easily yield the correct answers.

This method of producing question papers from a bank of questions can be organized as depicted in the Entity Relationship (E/R) diagram shown in Figure 6. The key differences between the previous E/R diagram and this one is the different versions of texts for questions, multiple versions of answers, and an individual paper for each student.
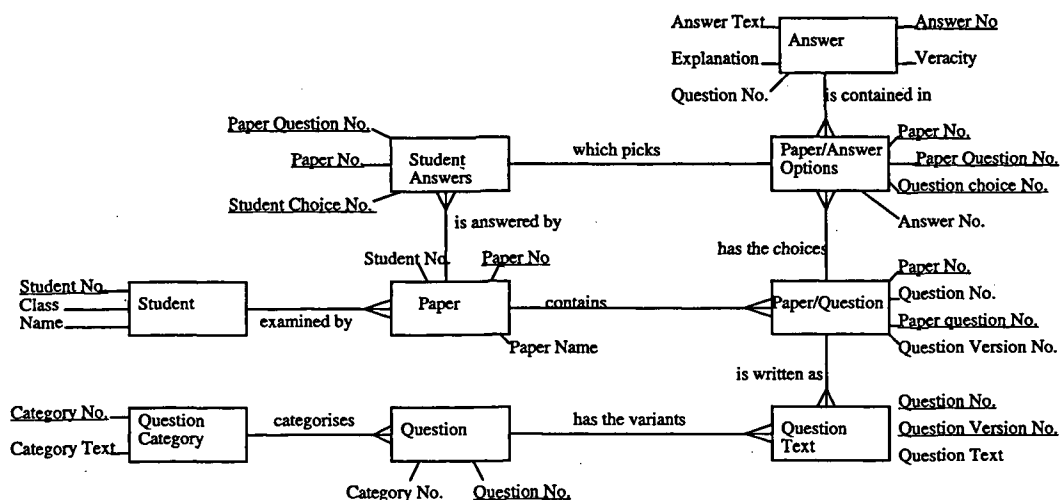
Figure 6: Entity Relationship diagram for individual multiple-choice question papers

Implementation

The software for this project was written in Borland C++ by P. Morgan (UWCC) using a modified version of the earlier E/R diagram which could be implemented more efficiently. Each entity was stored as a separate file. The modified E/R diagram is depicted in Figure 7 using the same notation as found in Figure 6. A breakdown of the menu structure can be found in Figure 8.

Figure 7: Modified version of an individual multiple-choice question paper

Figure 8: Menu structure for an implementation

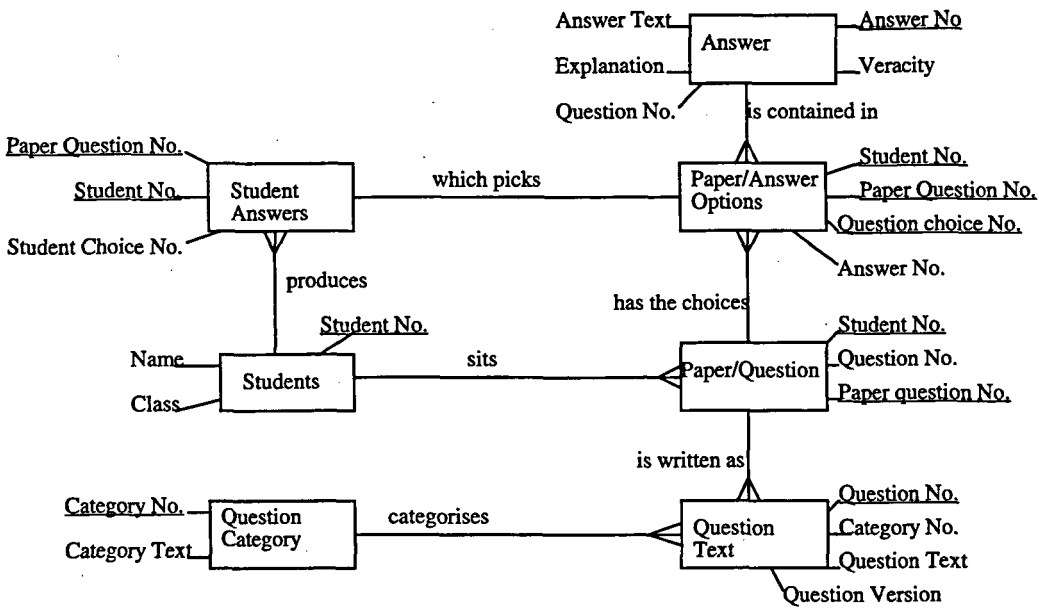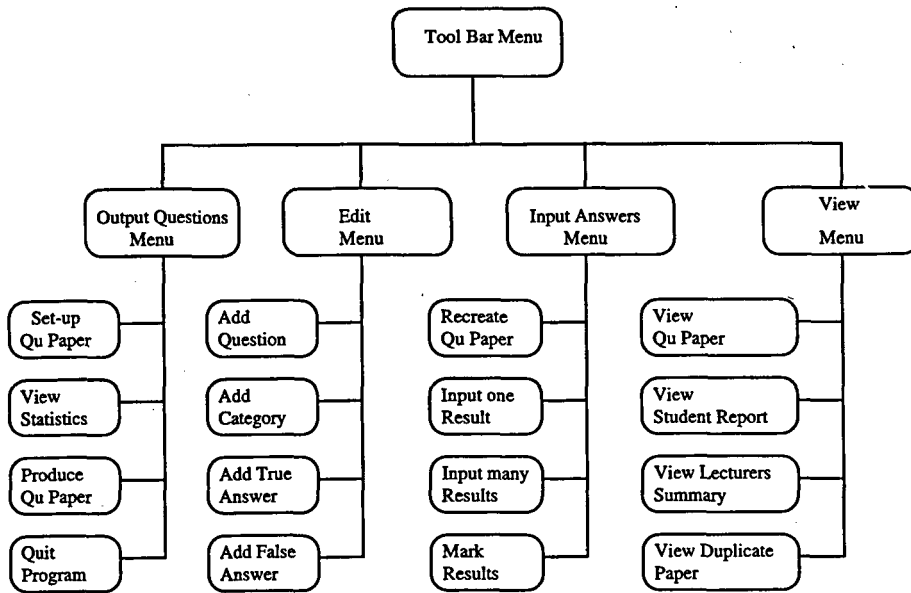In order to allow a high degree of flexibility in the system, the mark scheme can be selected from a number of alternatives. For example:

- 1 – Correct Answer   0 – False answer   0 – Don't know

- 1 – Correct Answer $-1/(c-1)$ – False answer   0 – Don't know
  where c is the number of choices for a multiple-choice answer.

- 1 – Correct Answer   −1 – False answer   0 – Don't know

In this paper the results are based on the second option, which is a negative marking scheme.

Results

One hundred and thirteen students were tested using ten questions from a bank of fifteen questions. Each question had the following choices: one question-text from two choices, one true answer from two choices, five false answers from seven choices and a Don't know option. Questions were allocated in a random order on the paper, and choices allocated randomly within a question. The probability of a number of questions (s) being duplicated when comparing two individual papers agreed with the analytic result given earlier (see Figure 9). The two options given in this figure relate to two separate ways of defining 'different' papers. If the definition for option (i), as described earlier in this paper, is used, then the probability of finding less than or equal to three shared questions when comparing two papers is over 90%. Alternatively, the definition in option (ii) would describe the chance of finding one or fewer shared questions between two papers as over 90%. Both definitions are reasonable, therefore both results are given. An example of the type of question produced for the student's questionnaire is given in Figure 10.

Figure 9: Probability of shared questions {$b=15$, $q=10$, option i) $v=t.n=4$, option ii) $v=t.(c-1)C_f n=84$}

```
******************************************************************************

  STUDENTS NAME : MR    XXXX XXXXX XXXX        H610

    CLASS NAME : LLL93 )

MARKING SCHEME : [Right = (+1.0)][wrong = (-0.2)][Dont Know = (0.0]

******************************************************************************


[Question (1)
Given the following machine code and the table below, state the number
placed by the program at location $42, given that memory location (0041) = 3Hex

  Program:(7F 00 40)(DE 40)(A6 50)(97 42)(3F)

    Table: Memory Address (Hex) | Entry (Hex) | Entry (Decimal)
                        0050     |    00       |    0
                        0051     |    01       |    1
                        0052     |    04       |    4
                        0053     |    09       |    9
                        0054     |    10       |   16
                        0055     |    19       |   25
                        0056     |    24       |   36
                        0057     |    31       |   49


Answer (1)   Output from program = (10) Hex
Answer (2)   Output from program = (31) Hex
Answer (3)   Output from the program is impossible to determine as it is
             outside the range of the table.
Answer (4)   Output from program = (36) Decimal
Answer (5)   Output from program = (34) Decimal
Answer (6)   Output from program = 1001 binary.
Answer (7)   (DONT KNOW)

******************************************************************************
```

Figure 10: Example question from the multiple-choice questionnaire

Six categories of question were available from which to generate the report. These categories were used to identify 'weak' areas for tutorial work. Finally, the results were correlated with those previously obtained from a traditional examination, and they produced a correlation coefficient of 0.923 for 111 degrees of freedom and two variables. This indicates a high degree of positive correlation between the results.

Feedback was also obtained from the students. They were not told how the questionnaires were generated, so the plagiarism question refers to their impression from their papers, not their understanding of the technique. An example of the questionnaire is given in Figure 11, with shaded boxes to indicate the percentage of students choosing the option indicated. Figure 12 shows the percentage results graphically.

In addition to the results shown in the figures, the following comments, derived from the comments section of the questionnaire, are worth highlighting.

• Students found that multiple-choice questions specifically designed with options that tested their ability to understand subtle differences and exploited common student misunderstandings were very difficult to answer. One commented that this forced them to go to the textbooks.

Multiple Choice Questionnaire (14/1/94)

The following questionnaire is designed to elicit your reaction to using multiple choice papers with mixed questions and computer generated explanations.

Please tick one box and place any relevant comments at the bottom of the page.

Did you find the multiple choice paper	Very Difficult	Difficult	Reasonable	Easy	Very Easy

Did you prefer the MCQ paper to traditional methods	Much Preferred	Preferred	Made little Difference	Disliked	Very much Disliked

In your opinion would Plagiarism be	Very Difficult	Difficult	Moderate	Easy	Very Easy

Did you think the Feedback was	Much better than normal	Better than normal	Average	Poor	Very Poor

Did you think that the speed of response (feedback time) was	Excellent	Good	Reasonable	Poor	Very Poor

Do you think as a teaching tool this form of MCQ was	Excellent	Good	Reasonable	Poor	Very Poor

Do you think as an assessment tool this form of MCQ was	Excellent	Good	Reasonable	Poor	Very Poor

COMMENTS

KEY   - indicates percentage of students ticking the option shown (50% uptake of questionnaire)

30%	25%	20%	10%	5%	0%

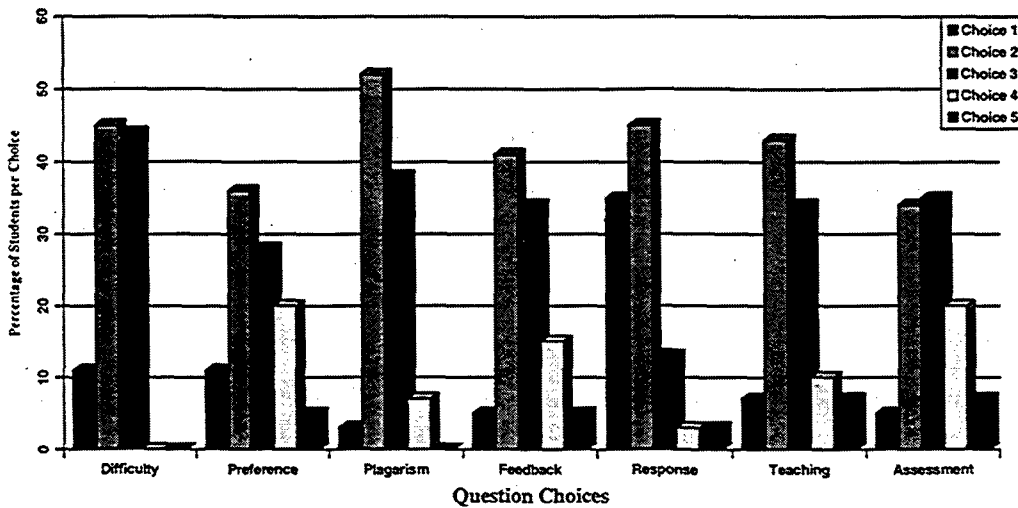*Figure 11: Student questionnaire*

Figure 12: Bar chart showing the percentage results for the questionnaire

- The assessment was criticized for using negative marking and not being able to give part-marks for working out. This indicates that the students' opinion on the multiple-choice questionnaire as an assessment tool may partly be based on their opinion of negative marking.

- The computer-generated answers only referred to one version of the question text. Some students found this confusing when trying to understand the feedback papers.

Conclusion

According to the questionnaire results, the class seemed to prefer this form of assessment despite finding it difficult. Plagiarism did not seem to be a problem, and the response time was good. However, they would have preferred a system that did not use negative marking, and would have liked clearer explanation sheets. Of particular interest is the students' opinion of this CAA as a teaching tool. Students believed this CAA system was a better teaching tool than assessment tool, according to the questionnaire. This statistic is backed up by comments on how the CAA paper forced them to study their books and which also indicated their dislike of an assessment method that used negative marking.

Clearly, this particular study achieved its goals and provides an effective means of assessing students and providing effective feedback. However, three man-days were spent on developing the questions by a lecturer experienced in setting appropriate questions for this subject. So no saving of time can be realized unless very large classes (>100) are assessed or the questions can be re-used. Due to the nature of this system, such questions could be re-used provided a student did not collect multiple explanation sheets from a previous year. Arguably, a student who reads through explanations for many different forms of questions and answers is likely to spend more time learning than the one who studies his or her course notes. Naturally, the explanations could be withheld, thus treating the students from several years as one class for the purposes of the results obtained. But this defeats the feedback aspects of the system. Certainly, this system does

not represent a perfect solution, as considerable effort has to be put into the 'false' answers in order properly to test and inform the students. Nor can this be used as a system to replace traditional assessment, as traditional assessment and tutorials provide the means for identifying common misconceptions. Nonetheless it is a powerful new tool in the hands of a lecturer. The next stage in developing this system will be to provide a networked version to remove unnecessary paperwork and to rewrite the system so that it becomes part of a standard database. These developments should allow greater flexibility and greater ease of use.

References

Committee of Scottish University Principals (1993), 'Teaching and learning in an expanding higher education system: executive summary', *The CTISS File*, 15, 5–6.

CVCP Briefing Note (1994), *Funding of Undergraduates and University Teaching*, London, CVCP Publications.

Benford, S., Burke, E., Foxley, E., Gutteridge, N. and Mohd Zin, A. (1994), 'Early experiences of computer-aided assessment when teaching computer programming', *Association for Learning Technology Journal*, 1 (2), 55–70.

Fry, J. (1990), 'The database format question: an alternative to multiple choice and free format for computer-based testing', *Computers and Education*, 14 (5), 391–401.

Laurillard, D., Swift, B. and Darby, J. (1994), 'Academics' use of courseware materials: a survey', *Association for Learning Technology Journal*, 1 (1), 4–15.