Assessing multiple choice question (MCQ) tests - a mathematical perspective

Scharf, Eric M.; Baldwin, Lynne P.

Mitglied der
Leibniz-Gemeinschaft

Assessing multiple choice question (MCQ) tests – a mathematical perspective

ERIC M. SCHARF   *Queen Mary, University of London, UK*

LYNNE P. BALDWIN   *Brunel University, UK*

ABSTRACT The reasoning behind popular methods for analysing the
raw data generated by multiple choice question (MCQ) tests is not
always appreciated, occasionally with disastrous results. This article dis-
cusses and analyses three options for processing the raw data produced
by MCQ tests. The article shows that one extreme option is not to
penalize a student for wrong answers or for missing out questions, and
the other extreme option is actually to penalize both aspects. The inter-
mediate option of focusing on the number of questions actually
attempted while penalizing wrong answers can be regarded as the
fairest. In this case blind guessing will on average not help the student,
although partial knowledge will lessen the negative impact on the final
overall score. There are still many interesting challenges in designing
techniques for MCQ tests.
KEYWORDS: *computer-based assessment, examinations,
multiple choice question tests, penalizing guessing, test scores*

Introduction

Assessment in higher education is always on the agenda, given its central-
ity to the awarding of degrees of one level or another. Every mark that a
student gets for an item of coursework or an examination makes its
contribution to the mark or grade for the module as a whole which, in
turn, affects degree classification in some way. For example, classified
undergraduate degrees in the UK are known as Honours degrees, and have
four 'categories': first, upper second (2.1), lower second (2.2) and third.

Like us all, students wish to do as well as they can, and to achieve the best qualification that they can. In today's climate, in the UK at least, most employers state in their job advertisements that potential candidates will normally need an upper second (2.1) and so there is pressure for students to achieve this. With higher education being viewed more of a service to be bought by students and with us in universities being viewed as those who deliver that service, it is not surprising that every mark counts. This can be made particularly challenging by the way that these degrees are calculated.

In order to recommend a candidate for a degree award, the overall weighted mark average has to be calculated. Precisely how the final mark is weighted varies from institution to institution, but it is not uncommon to base it on the average mark from level 2 and the average mark from level 3 and to weight level 2 at one third and level 3 at two thirds based on the 'best' marks achieved in 80/100 credits out of 120 credits at each level. The decision-making of the board of examiners is usually made easy if the overall weighted average falls well within a range. For example, if a mark of 65 falls between 60 and 69, the examination board would normally recommend an upper second degree. Naturally enough, if a candidate has had some personal circumstances that have affected their academic performance, the decision-making is not so straightforward. However, at some point during the meeting of the board of examiners, we are faced with a candidate who has an overall weighted mark average of, say, 69.87. Depending on the discipline, institution or practices of the board of examiners in question, this will or will not be considered as a 'borderline' case. Some boards of examiners may rightly claim that as 69.87 is not a mark between 70 and 100 (the mark required for the award of a first class degree), the appropriate award should be an upper second (2.1). For others, a weighted mark average of 69.87 is considered so near as to be 'pretty much' a mark of 70. Such decision-making is known as 'academic judgement', which is exercised differently in different institutions, or indeed disciplines, for various reasons. However academic judgement is exercised, such seemingly small differences in the mark average can make the difference between a higher or a lower award which, in today's climate, can in one way or another affect our students.

To ensure that we make the most appropriate award for each student, we ask students to demonstrate their achievements by way of various assessments. Some would argue that this assesses not what they *know*, but what they can *show* us, which may be two very different things (Race, 2005: 67). Many would probably agree with Race (2005) that 'assessment is the principal driving force for learning for so many learners'

in higher education today and that this goes against what we claim institutions of higher education are about, that is, promoting and fostering learning itself. Many would also probably agree with Race (2005: 66) that we, like our students, are driven by results, given that our funding depends on showing that this learning has happened and that we have played our part in this. Whether or not this is the case, assessment is certainly at the top of the agenda in learning and teaching in higher education institutions, whatever the complex reasons, and we spend enormous amounts of time on making assessment 'valid, reliable, transparent and authentic' (Race, 2005: 74). Indeed, some might argue that we spend far more time on assessment-related matters than we do on learning and teaching itself.

For those of us with large classes of 300 plus (a class of 100 is now regarded as 'small'), the choice of the means of assessment appropriate to a particular course needs to be carefully considered if we wish to devote more of our valuable time to interacting with our students. Ideally the assessment approaches should not be selected purely on the basis of ease or speed of marking. There are two main types of assessment; normative-referenced and criterion-referenced. In addition, there is also self-referenced (ipsative) assessment, which is gaining in popularity in the more student-centred (rather than teacher-centred) approaches that many claim to use in higher education (Lindblom-Ylanne et al, 2006; Pitts, 2005). Whether formative or summative in nature, assessment of, and feedback to, learners shapes how learners think and behave; this is further evidence of our need to take great care. Feedback plays a central role in self-regulation skills in learning, required when comparing performance against the goals that have been set (Nicol and Macfarlane-Dick, 2006). For a useful overview and in-depth discussion of assessment more generally, the reader is directed to Freeman and Lewis (1998), Brown and Glasner (1999) and Race (2005).

One of the 'tools' in our 'toolbox' of assessment methods are objective tests. These fall into three main types, namely true/false, matching and multiple-choice questions (MCQs), in the context of which, say Freeman and Lewis (1998: 145), we must be clear what we mean by 'objective'. It is, they say, that marking can be done mechanically (more often than not by a computer, these days) as no judgement needs to be made. To call such a test 'objective' does *not*, however, mean that MCQ tests and the like assess candidates 'objectively', that is, without any prejudice or bias. Indeed, such a thing is not possible. *No* assessment method can be objective; *all* assessment methods are subjective to a greater or lesser extent. As markers, we can be influenced by gender, ethnicity and the like, and knowing, or not knowing, the identity of

the students whose work we are marking (if students do or do not have their name on their work) affects the marks we award (Fleming, 1999, in Brown and Glasner, 1999). As Barrow (2006: 206) rightly says, 'all assessment places students in a web of power'; given this, we need to think long and hard about how we *use* that power.

There is a wide range of views on the effectiveness of MCQ tests. According to Freeman and Lewis (1998: 147), MCQ tests can test only facts and are of less substance than other forms of assessment; these authors go on to say that such tests have the reputation, undeserved, of being 'easier' than other methods and also of being able to test only 'lower level' skills and abilities. Whatever their limitations, Brown and Glasner (1999: 11) claim that MCQ tests are 'much more sophisticated than many of us once believed', citing their use by the Open University (OU) in the UK and directing readers to the OU for examples of best practice. Race (2005: 86), too, supports this view, adding that if we design MCQ tests well, we can be as confident as we can be with any assessment, that such tests will tell us the extent to which students have made sense of what they have learned and that we can ensure that our marking will not be affected by illegible handwriting or that students will not be penalized for writing slowly. Indeed, Race (2005) goes on to say that there are many more benefits from MCQ tests, among these that, by providing on-screen feedback immediately, students may avoid taking with them the errors from the earlier questions and that given the speed at which they get feedback, they can still remember *why* they thought and did what they thought and did, and can thus reflect on this. MCQ tests are, it seems, popular with learners although they may be perceived as being less fair (Struyven et al, 2005).

With the increasing use of computers to support our students' learning these days, many software packages now do much of the work for us in terms of helping us to design MCQ tests which we can use in our modules. We may ask students to carry these out in class, via, say, WebCT or Blackboard (most institutions in the UK now use one or other of these packages for, as a minimum, the 'basics' of the module content), or perhaps for self-assessment outside of the classroom (Blackboard, WebCT). Again, with the relative sophistication of the software these days, it is not too difficult – with, perhaps, a little help for those who are not so technically experienced – to set up such tests.

Naturally, the course organizer needs to build up a sufficiently large bank of questions so that even if students see these questions before the test, the large number of questions effectively ensures novelty each time students take the test.
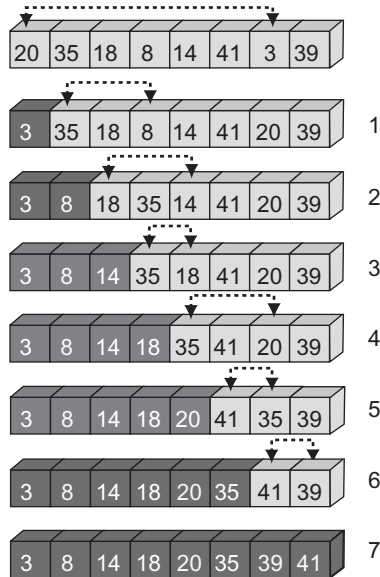
Figure 1 **Typical multiple choice question on sorting algorithms**

The dark items with white lettering indicate the sorted items; the lighter items with black lettering indicate the unsorted items. The numbers 1 to 7 on the right-hand side indicate the particular stage of the sorting process.

Which sorting algorithm is represented by the figure?

A partition sort   C insertion sort
B bubble sort   D selection sort

MCQ tests, if appropriately designed, test not only the ability of students to memorize and recall knowledge but also test the ability to reason analytically with that knowledge. The example in Figure 1 shows how this is possible. Here the student is asked to determine the type of sorting algorithm that the diagram represents. This not only requires memorization of facts but the ability to appreciate how the relevant sorting algorithms work and to compare that knowledge with the information presented by the diagram. Questions on a segment of program code provide further examples of how MCQ tests can test not just powers of information retention and retrieval but also those of analysing and reasoning with knowledge in situations where context is important.

Admittedly, MCQ tests, unlike written examination papers which call for longer responses, do not show the steps taken by students to reach their answers. In addition, MCQ tests may not test the ability of students to create or synthesize new knowledge from previous items that have been learned

and understood. It is for these reasons and others that MCQ tests are but one method of assessment and that we, as course organizers/assessors, need to exercise care to ensure that we are testing that which we wish to test. For example, a course in computer algorithms could include an MCQ test to examine students' knowledge of the subject (for example, linked list manipulation) and to exercise the ability to reason with that knowledge. However, such a course should also include exercises that enable a student to apply that knowledge to solve particular tasks (for example, application of linked lists to symbolic algebraic manipulation such as addition, multiplication or integration). Nonetheless, the nature of MCQ tests make them ideal as an assessment tool for first year university courses, where the nature of the material and large class sizes can really bring out the benefits.

Whilst the benefits and drawbacks of MCQ tests were discussed earlier in this article, the basic issues of how we can convert the raw data generated by an MCQ test into meaningful results are rarely raised. It is important to create a model that gets as close as possible to a student's reaction to, and processing of, the questions in an MCQ test, while not unduly penalizing the student. In addition, to what extent does guessing benefit the student at the expense of a fair assessment by an MCQ test of that student's abilities? The use of MCQ tests as an assessment tool is common in the higher education sector and is growing in popularity. It is argued here that course organizers/assessors who compile these tests in the educational environment may on occasion make serious errors when interpreting the raw data generated by MCQ tests if not aware of the issues raised within this article. These factors suggest that there is a need to consider the more formal mathematical aspects of this important topic if we are to ensure that the marks that we award are appropriate.

The basic equation

Let R be number of questions for which a student gives a right answer and K be the number of questions for which a student actually knows the right answer. *The vital point is that R may be greater than K.* This would typically be the case if the student has actually guessed the answers to the additional $(R-K)$ questions, because of lack of knowledge or application of appropriate reasoning. (Note, as a help to the reader, the symbols used in the equations in this and the following sections are brought together in the appendix at the end of this article.)

Assume that all questions carry equal weight and that for each question there are A possibilities. In addition, let Q be the total number of questions attempted out of an overall total of T questions. Now $((Q-K)/A)$ will be the number of questions that the student has — on average — got right by

chance, since there are *A* possibilities for each question. Hence, the number (R) of questions for which a student *gives* a right answer is related to the number (K) of questions for which the student *knows* the right answer, by the following equation:

$$R \;=\; K \;+\; \frac{Q \;-\; K}{A} \tag{1}$$

or, rearranging,

$$K \;=\; R \;-\; \frac{Q \;-\; R}{A \;-\; 1} \tag{2}$$

The second equation (2) shows that the $(Q-R)$ wrong results will penalize the student. Note, additionally, that in this equation, negative values of K can be forced to a floor value of zero, which is then the minimum mark a student can achieve.

Three options for multiple choice tests

Three options with differing approaches to penalizing guessing can be identified. These are the zero, intermediate and maximum penalty options.

Option 1 – zero penalty

Here the input marks, apart from scaling, can be regarded as raw data and are not subject to further processing (that is, they are not subject to any reduction).

In this case students can gain marks by guessing (that is, by luck), and, on average, $1/A$ of the guesses will be correct, *even if the students have no knowledge of the actual topic being examined*. Wrong results are not penalized. Thus in equation (1), if $K=0$ (i.e. the student has no knowledge of the subject in which they are being assessed), the number R of questions for which the student gives the right answer, will still, on average, be (Q/A). This simple approach does not attempt to distinguish between answers resulting from genuine knowledge, and those resulting from lucky guesses. This scenario is different from a conventional written paper, where to gain any marks at all, some correct response – written or pictorial – is normally required of the student.

A typical exam rubric for this version of an MCQ test could read as follows:

> Each question is provided with 4 answers, only one of which is correct. Answers are identified by the letters A, B, C and D. Work out which letter corresponds to the best answer and put its letter legibly in the box provided. If

the letter is not legible, then no mark will be given. THERE IS NO PENALTY FOR A WRONG ANSWER. YOU ARE THEREFORE URGED TO ATTEMPT EVERY QUESTION. If you feel you do not know the answer to a question you should select the answer that you feel is the most likely.

The zero penalty MCQ test option is, of course, suitable, and very popular, when MCTs are used in a self-test (didactic) mode, in which the student is learning, as opposed to being assessed.

Option 2 – penalty based on questions attempted

In this case, the number W of questions the student answers wrongly is given by:

$$W = Q - R \qquad (3)$$

Equation (2) can be used to find K, the number of questions correctly reflecting the student's knowledge, in terms of W. This gives equation (4):

$$K = R - \frac{W}{A - 1} \qquad (4)$$

This (intermediate) option is widely used when assessing MCQ tests, since while it penalizes wrong answers, it takes no account of the number of questions which the student has not attempted.

   With this intermediate option, a student attempting a given number X of questions, all of which are answered correctly, may get more marks than their colleague who attempts Y questions, where $Y > X$, but where still only X questions are answered correctly. The justification for this is that the 'Y' student has used more chances. This is emphasized in the table and the graph in Figures 2 and 3, respectively. Both consider a scenario of 20 questions, with 4 choices per question – one choice of which is correct. The number (R) of correct questions has been multiplied by 5 to give a percentage mark.

   When using the intermediate option 2, a possible rubric for an examination paper could read:

> You will only be assessed on questions which you have attempted and to which you have provided a legible answer. It is recommended that you avoid guessing an answer to a question.

Option 3 – maximum penalty

A student can be further penalized by option 3 (maximum penalty) which not only considers the number of questions that were *incorrectly answered*, as in option 2 above, but also considers the number of questions that were *not attempted*. Thus questions not attempted are also assumed to be incorrectly answered.

In equation (2), replace Q by T, the total number of offered questions, and K now becomes K_M, the number of questions that it is assumed have been correctly answered as a result of the student's knowledge.

$$K_M = R - \frac{T - R}{A - 1} \tag{5}$$

This equation also represents the limiting case of equation (2), when all the questions have been attempted.

Another form of the Maximum Penalty Equation can be derived, by assuming M marks per question, and putting $S = K_M M$ and $S_{MAX} = TM$. This gives the following equation:

$$\frac{K_M}{T} = \frac{S - \left(S_{MAX}/A\right)}{S_{MAX} - \left(S_{MAX}/A\right)} \tag{6}$$

Numerical and graphical comparisons

Table 1 considers a scenario of 20 questions, with 4 choices per question – one choice of which is correct. For convenience of presenting the numerical results in this article, the number of correct questions has been scaled by 5 to give a percentage mark. Both columns 1 and 2 can thus be regarded effectively as the same raw data but with different scaling. Note that usually one tries to avoid negative marks, so that K has not been allowed to go negative. The three options are represented as follows.

Option 1. With option 1 (zero penalty), there is no further processing as column 2 of Table 1 shows.

Option 2. With the intermediate option (option 2), a student attempting a given number X of questions, all of which are answered correctly, may get more marks than their colleague who attempts Y questions, where $Y > X$, but where still only X questions are answered correctly. The justification for this is that the 'Y' student has used more chances. This is emphasized in Table 1 and Figure 2, respectively. For example, for the actual score of 8 questions answered correctly (but possibly including guesses) or 40 per cent, the intermediate approach gives marks of 33 per cent, 25 per cent and 20 per cent for $Q = 12$, 17 and 20, respectively.

Option 3. The fourth column also represents option 3 (maximum penalty) as the limiting case of option 2 ($Q = T = 20$), where all offered questions have been attempted.

Table 1 **Comparing zero, intermediate and maximum penalty approaches**

Option Actual score = R	1. Zero penalty K = R Score (%) = 100*(R/T)	2. Intermediate Q = 12 Score (%) = 100*(K/T)	2. Intermediate Q = 17 Score (%) = 100*(K/T)	3. Max. penalty Q = T = 20 Score (%) = 100*(K/T)
00	0	0	0	0
01	5	0	0	0
02	10	0	0	0
03	15	0	0	0
04	20	7	0	0
05	25	13	5	0
06	30	20	12	7
07	35	27	18	13
08	40	33	25	20
09	45	40	32	27
10	50	47	38	33
11	55	53	45	40
12	60	60	52	47
13	65	–	58	53
14	70	–	65	60
15	75	–	72	67
16	80	–	78	73
17	85	–	85	80
18	90	–	–	87
19	95	–	–	93
20	100	–	–	100

Total no. of questions = Q = T = 20.
Possible no. of answers per question = A = 4.

Figure 2 also shows that the operation area for MCQ tests is bounded by the triangular area formed by the zero and maximum penalty lines and the x-axis.

The influence of processing an MCQ test on a student's mark

The graph in Figure 3 is derived from the graph in Figure 2 and shows the differences in the scaled output mark (K) (that is, reading in the vertical direction) between the zero penalty approach and intermediate penalty approaches where the number of questions attempted is 4, 8, 12 and 16, respectively, with the limiting case of $Q = T = 20$ being equivalent to the maximum penalty approach. The maximum difference is as much as 25 per cent.
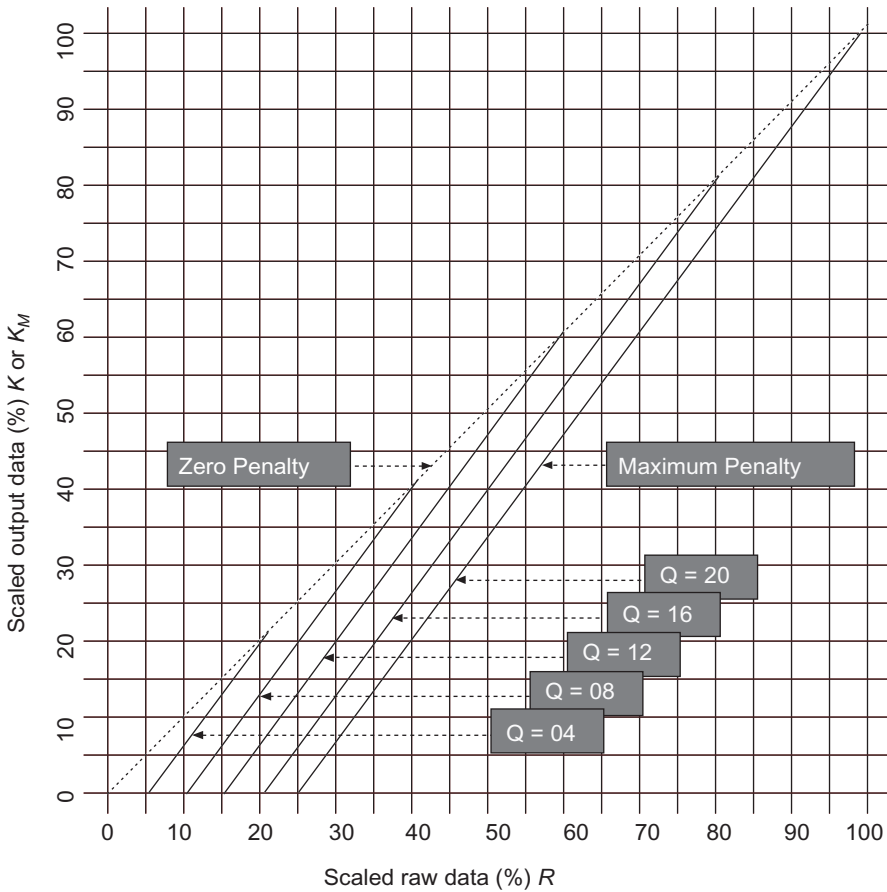
Figure 2 **Graph comparing zero, intermediate and maximum penalty approaches**

Figure 3 provides a stark illustration of how the mark that a student obtains for an MCQ test is strongly influenced by the way in which the educator processes the raw marks generated by the MCQ test.

Does the student benefit by guessing?

With Option 1, the answer is definitely yes for 'blind' guessing. However, on average the answer is no if the intermediate option 2 is used. This is borne out by equation (2) and Table 1. If Q, the total number of questions attempted, increases, but R, the actual score, stays the same, then the mark K actually given to the student will decrease.

This is also summarized in Figure 4, which shows the number of questions attempted against the resultant mark K for different raw scores (R) where
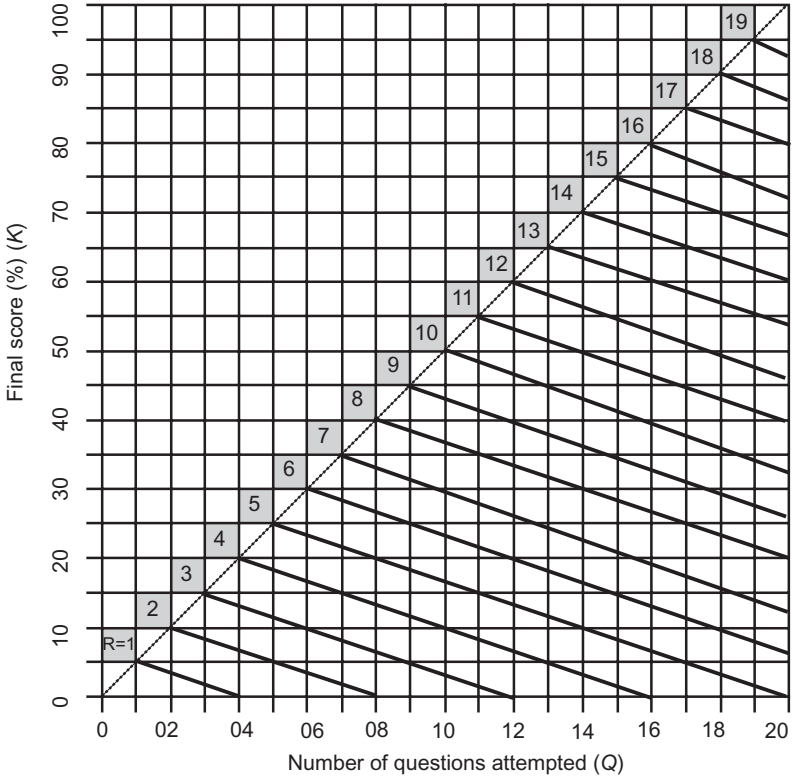
41

Figure 3 **Graph comparing differences between zero penalty and other approaches**
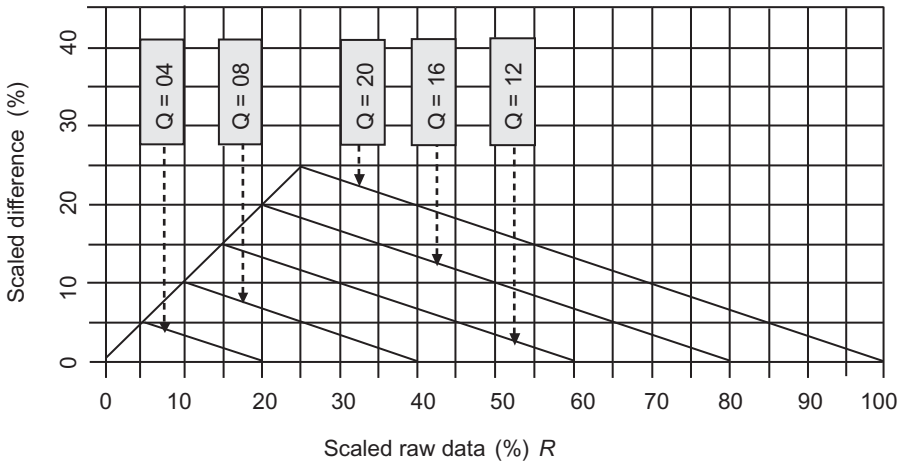

Figure 4 **Number (Q) of questions attempted against final score (K) for different numbers (R) of questions for which the correct answer is given (either from knowledge or by chance)**

R can be regarded as the number of questions apparently correct. Note that R is expressed in terms of the number of questions answered correctly, irrespective of whether this is due to the student's actual knowledge or to guessing. The downward sloping lines for given values of R indicate diminishing returns for the student if wrong answers are given for further questions $(Q-R)$ attempted.

Suppose a student has additional partial knowledge that improves the chances of an answer being correct from $1/A$ to $1/B$ (i.e. $B<A$). For example, students may be able to narrow down their options from 1 in 4 choices ($A = 4$) to 1 in 2 choices ($B = 2$). In this case, in equation (4), the subtracted term ($W/(A-1)$) will be less in value if A is replaced by B, where $B<A$. With partial knowledge, the negative impact on the final mark will, on average, be lessened. Naturally, this is only a very crude model of the situation, since students will not normally approach each question with the same degree of partial knowledge.

Guessing and the number of question choices

Does the effect of guessing depend on the number (A) of possible choices per question? So far we have assumed for each question, four choices (i.e. $A = 4$) one of which is correct. From Figure 2 it can be seen that the maximum difference D_{MAX} in marks between the zero and maximum penalty lines is given by vertical distance between the two lines from the crossing of the x-axis by the maximum penalty line. For $A = 4$, this is 0.25 p.u. (per unit) or 25 per cent. This exercise can be repeated by re-plotting Figure 2 for different values of A (e.g. $A = 2$ to $A = 10$).

More generally, we can take equation (5), and put $K_M = 0$ and $R = D_{MAX}$. This gives equation (7):

$$D_{MAX} = \frac{T}{A} \tag{7}$$

From this we can see that the effect of accounting for guessing, when using the intermediate option 2, becomes less as the number (A) of choices per question increases. Essentially, as the number of choices increases, the probability of picking the correct answer by chance diminishes. The graph and table in Figure 5 summarize this.

What is the optimum number (A) of choices for each question? Commonly, but not exclusively, three or four choices are used. The greater the number (A) of choices, the more time must be allocated to a student to complete a question, but the smaller will be the influence of blind guessing by the student. From the educator's perspective, the effort
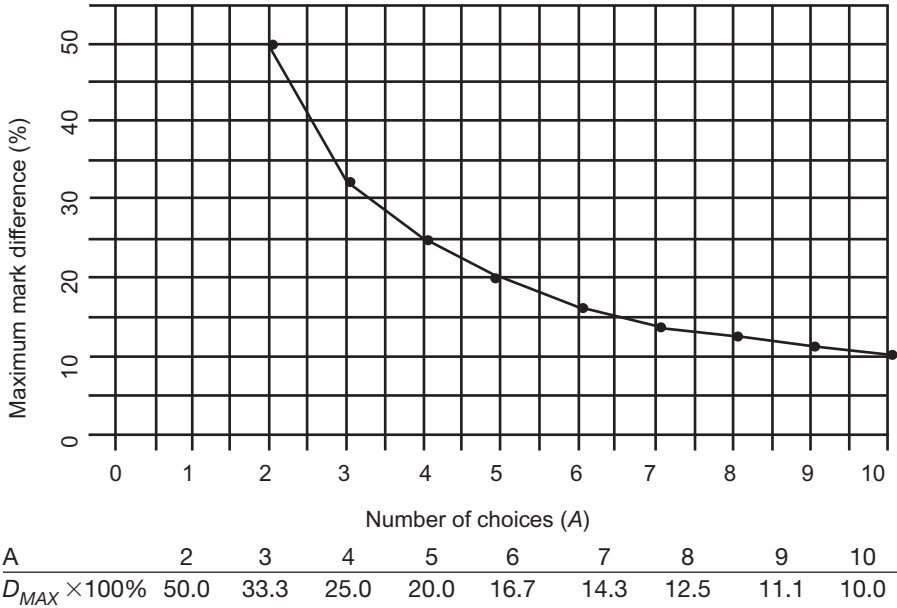
A	2	3	4	5	6	7	8	9	10
$D_{MAX} \times 100\%$	50.0	33.3	25.0	20.0	16.7	14.3	12.5	11.1	10.0

Figure 5 **Graph and table relating number (A) of choices to maximum possible absolute difference (D_{MAX}) in marks caused by accounting for guessing**

involved in providing a large number of choices per question means more time spent on assessment at the expense of contact time with the students.

In the context of option 1, consider the following weighted cost function F, for which the term (αA) indicates the time spent preparing the question, where α is a weighting factor and the term (β/A) indicates the effect of guessing, where β is a further weighting factor.

$$F = \alpha A + \frac{\beta}{A} \qquad (8)$$

If we assume that all variables and parameters in equation (8) are positive, then equation (8) has a minimum for:

$$A = \sqrt{\frac{\beta}{\alpha}} \qquad (9)$$

Thus, if minimizing the effect of guessing is regarded as being 9 or 16 times more important than the initial work in preparing the question, the corresponding number (A) of choices should be 3 or 4. This fits in with normal practice. The emphasis on 'initial work' highlights one advantage

for the teacher of building up and maintaining a sizeable bank of multiple choice questions. Of course, as has already been mentioned, another advantage of such a question bank is to ensure the effect of novelty in the questions for the student.

Future work

MCQ tests are very popular with students, especially when the tests are computer-based and they can thus receive feedback promptly. They are also popular with educators who, by saving on the time taken with marking, providing feedback and all the other tasks associated with assessment, can devote more resources to the actual process of teaching. The following issues show that there are still many challenges with setting and devising MCQ tests, however.

Revealing thinking patterns. Basic MCQ tests are said to test only factual knowledge, but by careful attention to the question, thinking processes could be deduced. Marks could then be allocated not only on the basis of whether the question was correctly or incorrectly answered, but also on how the student reached their conclusion. More work is needed here to formalize the process.

Reassuring the student – good interfaces. A good computer interface can encourage and put students at their ease. A good layout of a printed version of an MCQ test can also be beneficial to the student.

Reassuring the student – adjusting marking scheme. There are many possibilities. A basic one is to consider a compromise between the no-penalty option 1, which encourages students to supply answers, and option 2, which discourages guessing. This could be achieved by making the factor A less than the number of available question choices.

Conclusions

For a written paper, a student must have some knowledge of the subject, although a chance element still exists regarding exactly what that student actually chooses to put as the answer. However, the mark that a student obtains for an MCQ test is strongly influenced by the way in which the educator processes the raw marks generated.

Approaches to processing and analysing the raw data generated by MCQ tests that are used for assessment can be distinguished by the way the approaches deal with guessing by the student. One extreme imposes no penalty for wrong answers; in this case, even if the student has no knowledge of the subject being tested, they will get on average a mark of $(100/A)$ per cent, where A is the number of choices per question. The other extreme

penalizes both wrong answers *and* the lack of answers to questions and may be viewed as the option for which there is the least justification.

If we are to compensate for guessing, then the intermediate option will give more marks to each student than the maximum penalty option. The intermediate option, on average, penalizes blind guessing, although a partial knowledge lessens the negative impact on a student's final mark. With this option, the effect on a student's marks of accounting for guessing also becomes less as the number of choices per question increases, giving, for example, maximum absolute differences in marks of 33 per cent and 25 per cent for three and four question choices, respectively.

All three approaches obviously need to know the number (R) of right answers and the marks per question. In addition, the intermediate and maximum penalty options need to know the number (Q) of questions attempted and the total number (T) of questions offered in the MCQ test, respectively.

Multiple choice tests have gained popularity with students and educators, but there is considerable scope for further research, for example in techniques for reassuring and encouraging the student and techniques for revealing their thinking patterns.

Acknowledgements

References

BARROW, M. (2006) 'Assessment and Student Transformation: Linking Character and Intellect', *Studies in Higher Education* 31(3): 357–72.

BLACKBOARD, THE BLACKBOARD ACADEMIC SUITE™ (2006): http://www.blackboard.com [accessed June 2006].

BROWN, S. & GLASNER, A., eds (1999) *Assessment Matters in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.

FREEMAN, R. & LEWIS, R. (1998) *Planning and Implementing Assessment*. London: Kogan Page.

LINDBLOM-YLANNE, S, PIHLAJAMAKI, H. & KOTKAS, T. (2006) 'Self-, Peer- and Teacher-assessment of Student Essays', *Active Learning in Higher Education* 7(1): 51–62.

NICHOL, D. J. & MACFARLANE-DICK, D. (2006) 'Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice', *Studies in Higher Education* 31(2): 199–218.

PITTS, S. E. (2005) 'Testing, Testing …'; How Do Students Use Written Feedback?', *Active Learning in Higher Education* 6(3): 218–29.

RACE, P. (2005) *Making Learning Happen*. London: SAGE.

STRUYVEN, K. DOCHY, F. & JANSSENS, S. (2005) 'Students' Perceptions about Evaluation and Assessment in Higher Education: A Review', *Assessment and Evaluation in Higher Education* 30(4): 325–41.

WEBCT, WEB COURSE TOOLS: http://www.webct.com/ [accessed June 2006].

Appendix. Symbols and abbreviations used in this article

A	number of possibilities per MCT question
B	inverse probability of partial knowledge of a question
D_{MAX}	maximum difference in marks
F	weighted cost function
K	number of questions for which a student knows the right answer
K_M	number of questions for which a student knows the right answer (Maximum Penalty)
M	marks per question
Q	total number of MCT questions attempted by student
R	number of questions for which a student gives a right answer
S	product of K_M and M
S_{MAX}	product of T and M
T	total number of questions presented to a student in a MCT
W	number of questions wrongly answered
α, β	cost function factors

MCQ	Multiple Choice Question
MCT	Multiple Choice Test
p.u.	per unit

Biographical notes

ERIC SCHARF is a lecturer in the Department of Electronic Engineering, Queen Mary, University of London. He has been teaching topics dealing with software and microprocessors. He is an enthusiastic user of Multiple Choice Tests and has found that this enthusiasm is reflected in the positive approach taken by his students. Eric is research-active: he has successfully bid for, and led, a number of European Union-funded research projects including, more recently, TORRENT (concerned with applying intelligent agent technology to foster consumer choice in residential access networks) and CoRAS (a platform for the risk assessment of security-critical systems).

Address: Department of Electronic Engineering, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. [email: e.m.scharf@elec.qmul.ac.uk]

LYNNE P. BALDWIN is a senior lecturer in the Department of Information Systems and Computing at Brunel University, UK. She gained her PhD at Brunel University, and an MA in Language and Communication from the University of East Anglia, UK. Her research interests within the field of education and information systems are varied, although there is a strong emphasis on knowledge management, decision-making and related communication issues.

Address: Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. [email: Lynne.Baldwin@brunel.ac.uk]