

ALTERNATIVE MARKING SCHEMES FOR ON-LINE MULTIPLE-CHOICE TESTS

Martin Bush

School of Computing,
Information Systems & Mathematics,
South Bank University,
Borough Road,
London SE1 0AA, UK.
bushm@sbu.ac.uk
<http://www.sbu.ac.uk/~bushm>

ABSTRACT

Web technology lends itself very well to the provision of on-line multiple-choice tests, and many such systems are now in use. Many (if not most) of these incorporate the "traditional" marking scheme, with no adjustment to counter the effects of lucky guesses, in spite of the artificially high marks this gives rise to. A number of improved marking schemes have been proposed for written multiple-choice tests, some of which allow, or require, candidates to do more than just select one answer per question. They are all suitable for use in on-line testing. This paper reviews the most promising of these schemes, any one of which would (arguably) be an improvement over the traditional one.

Keywords

Multiple-Choice Testing, Web-Based Assessment, On-Line Testing, Education.

1. INTRODUCTION

Multiple-Choice (MC) tests are a widely used and attractive assessment technique. They can be used to objectively assess cognitive, analytical and other such skills in addition to straightforward factual knowledge. Web technology lends itself very well to the provision of on-line multiple tests, and many such systems are now in use. They are a perfectly respectable form of assessment; for example, Microsoft's well-known "Certified Professional" qualifications have been based purely on on-line MC tests for years.

The "traditional" marking scheme – one mark for every correctly selected answer – has the major drawback that candidates generally achieve artificially high marks due to lucky guesses. However, it seems that this fact is often ignored. It is not difficult to calculate the effects of lucky guesses.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

7th Annual Conference on the Teaching of Computing, Belfast

© 1999 CTI Computing

For example, to achieve a mark of 40% (on average) in a test consisting of four-answer questions, a candidate only needs to know the correct answer to 20% of the questions; they will get a quarter of the remaining 80% right (on average) through sheer guesswork. In reality candidates should be able to make informed rather than blind guesses, but this analysis at least demonstrates the point.

On-line tests offer two distinct advantages over paper-based tests. One is that they can deliver randomised variations of a test so that candidates are presented with a different version on each occasion [2]. Another is that they can provide immediate feedback, which potentially means that candidates can have repeated attempts at answering questions which they have so far answered incorrectly, although clearly this demands a non-traditional marking scheme.

A number of alternative marking schemes have been proposed (and used) over the years for written tests which are an improvement over the traditional scheme. In some cases these allow, or require, candidates to do more than just select one answer per question. They are all suitable for on-line testing.

2. ALTERNATIVE MARKING SCHEMES FOR WRITTEN MC TESTS

2.1 Permutational Questions

The likelihood of gaining marks through sheer guesswork can be drastically reduced by having questions with multiple correct answers, where the candidate has to select the correct permutation of answers to get the question right. Faced with a four-answer question for example, a candidate would have to choose correctly between 16 (i.e. 2^4) alternative responses, assuming the number of correct answers is unknown. This is a rather exacting test method, but the marking scheme can be relaxed in some way to reward students whose selections are nearly correct. One such scheme has recently been on trial at the University of Glamorgan [3].

2.2 Order-of-Preference Schemes

In an order-of-preference scheme there is only one correct answer per question. Candidates are required to assign an order of preference to all the answers for each question; their score is dictated by the preference they assign to the correct answer. For example, 4 marks could be awarded if the correct answer is the candidate's first choice, 2 if it is their second choice, 1 if it is their third. Order-of-preference schemes have been investigated by several authors [5].

2.3 Confidence Assessment

MC tests incorporating confidence assessment require candidates to assign a confidence level to each of their selections to reflect their degree of certainty. One such scheme, used on medical students at University College London [4], requires candidates to attach a confidence level of 1, 2 or 3 to their selected answer for each question; this is the mark awarded if their selection is correct, while 0, -2 or -6 are awarded (respectively) otherwise. Given the nature of the medical profession, it does seem particularly appropriate to penalise medical students whenever they give a confident but incorrect diagnosis!

2.4 Normalisation and Negative Marking

If a traditional one-tick-per-question MC test is preferred, some form of mark adjustment can be applied to counteract the effect of lucky guesses. One approach is normalisation, as depicted in Figure 1. For a test with four-answer questions the formula $x \rightarrow (x-25)*4/3$ (where x is the unadjusted test mark) would be used.

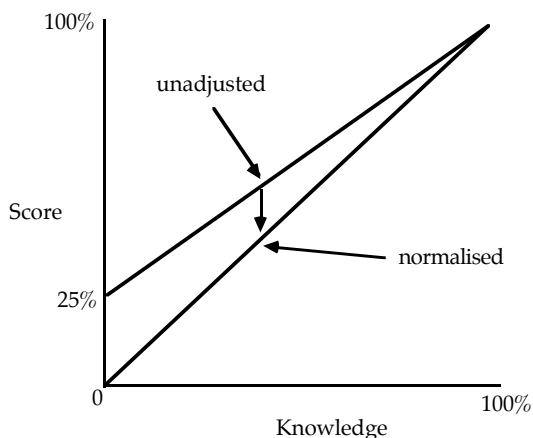


Figure 1. Normalisation of marks (assuming four answers per question).

An alternative approach is to use negative marking as follows; award $n-1$ marks for every correct selection (where n is the number of answers per question) but subtract one mark for every incorrect selection. This has roughly the same effect as normalisation, although it can result in a negative

overall mark if a candidate is unlucky. Both of these approaches are well known and quite widely used.

2.5 Liberal MC Tests

The use of negative marking opens the door to allowing candidates to select more than one answer to a question if they are uncertain which is the correct answer. This novel marking scheme is currently used on computing students at South Bank University [1].

To appreciate its implications consider the following, which assumes four-answer questions:

1. If a candidate knows the correct answer to a question, s/he gets $3/3 = 100\%$ for that question.
2. If the candidate knows that the correct answer is one of two options, s/he gets $(3-1)/3 = 67\%$ for that question, compared with an equal chance of getting either 0 or 100% in a standard MC test.
3. If the candidate knows that the correct answer is one of three options, s/he gets $(3-2)/3 = 33\%$ for that question, compared with having a 33% chance of getting 100% in a standard MC test.

The probabilities assumed here may not always reflect the true situation, because a candidate may not have equal faith in each of the answers s/he has chosen to select. Nevertheless it seems reasonable to suppose that the mark achieved in a liberal MC test is more likely to be a better indicator of a candidate's true knowledge than the mark that same candidate would have achieved faced with the same set of questions in a conventional test (even after normalisation). This is because candidates are able to express partial knowledge explicitly in a liberal test. In effect, candidates are forced to choose between a much richer set of alternative responses.

3. MARKING SCHEMES FOR ON-LINE MC TESTS

With on-line testing comes the potential for interactivity. A fully interactive MC test would be one in which candidates receive immediate feedback as soon as they answer a question. The most obvious marking scheme in this case is probably an "answer until correct" scheme [5], which corresponds to the order-of-preference scheme described above for written MC tests. Given that the marking is automatic it might be tempting to consider an even more sophisticated scheme than this, but it is essential to ensure that the assessment method is easily understandable and not a distraction to candidates.

Submitting each answer in turn and waiting for the result can be a frustrating experience if there is any waiting time involved. If it is preferred that candidates should answer all questions before submitting their answers, then the on-line test is just like a written test and any of the marking schemes described above can be used.

A sensible compromise might be to allow candidates a limited number of re-attempts at those questions they have so far answered incorrectly, with reduced marks for getting them right on subsequent attempts.

4. CONCLUSION

There are many valid objections that may be made against any kind of MC test, such as the difficulty of designing good questions, the possibility that thoughtful candidates may in certain cases read more into a question than the designer had in mind, and the fact that lots of questions are required to iron out the effects of lucky/unlucky guesses. On the other hand, MC tests have many very important benefits, which of course is why they are in such widespread use.

One objection to non-traditional marking schemes has been that the marking process is more error-prone [5], however this is not valid for on-line tests. Another is that unfamiliar marking schemes can be distracting to candidates at first [1], but this is only a transitional problem. A number of authors have reported that candidates do come to appreciate a sensible marking scheme once they become familiar with it, and that has been our experience at South Bank University.

Designers of systems for on-line testing should be aware of the alternative marking schemes that are available. Simple normalisation or negative marking can be used to counter the effects of lucky guesses without changing the nature of the test itself. Other marking schemes require candidates to do more

than just tick one answer per question, and can be used to explicitly assess confidence or partial knowledge. Designers should consider taking the opportunity that automation provides to move away from the traditional marking scheme to a richer, more sophisticated one.

5. ACKNOWLEDGEMENT

Many thanks to Steve Lerman for his encouragement.

6. REFERENCES

- [1] Bush, M. Allowing for uncertainty in multiple-choice tests. South Bank University Technical Report SBU-CISM-99-01 (1999) – available from the author.
- [2] Culwin, F. Web Hosted Assessment – Possibilities and Policy. Proc. of 3rd Annual Conference on Integrating Technology into Computer Science Education (ITiCSE' 98).
- [3] Farthing, D.W. (1998) Permutational multiple-choice questions: An objective and efficient alternative to essay-type examination questions. Proc. of 3rd Annual Conference on Integrating Technology into Computer Science Education (ITiCSE' 98).
- [4] Gardner-Medwin, A.R. Confidence assessment in the teaching of basic science. Association for Learning Technology Journal, 3, pp. 80-85 (1995).
- [5] Wood, R. Assessment and testing: a survey of research. Cambridge University Press (1991).