Practical Assessment, Research, and Evaluation

Volume 11 *Volume 11, 2006* Article 8

2006

A Comparison of Conventional and Liberal (Free-Choice) Multiple-Choice Tests

Sylvia Jennings

Martin Bush

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 11 Number 8, October 2006 ISSN 1531-7714

A Comparison of Conventional and Liberal (Free-Choice) Multiple-Choice Tests

Sylvia Jennings & Martin Bush
London South Bank University

We compare conventional multiple-choice tests with so-called "liberal" multiple-choice tests, also known as "free-choice" tests, from a theoretical standpoint. The style of the questions is identical in these two alternative test formats, but in a liberal/free-choice test candidates may select as many options per question as they wish; the marking scheme penalises incorrect selections via negative marking, to the extent that candidates have nothing to gain through blind guesswork. We show that in the absence of blind guesswork candidates really do get the marks they deserve in a liberal/free-choice test, since the format of the test does not introduce any statistical distribution whatsoever. This is the case even when the candidates have partial knowledge and can therefore engage in educated guesswork. By contrast, in conventional multiple-choice tests candidates will engage in guesswork whenever they are unsure of the correct answer. We also show that liberal/free-choice tests reward partial knowledge more generously than conventional tests do, while on the other hand they punish misinformed students more severely than conventional tests do.

There is a wealth of literature published on the subject of Multiple-Choice (MC) tests. The conventional MC test is one in which each question consists of a stem and C choices; one correct answer and C-1 distracters (incorrect answers). Candidates are told to pick one of the C choices. One mark is awarded for a correctly chosen option while zero marks are awarded for an incorrectly chosen one. Thus a candidate who knows the correct answer to K questions is assured of scoring at least K marks, as some of the other answers may be guessed. Unlucky guesses are not penalised and so it is in the candidate's interest to make a guess whenever the answer is unknown. This test format is commonly referred to as "number-right scoring" in the literature. An obvious disadvantage is that the resulting marks are likely to be inflated. In a test consisting of N questions, the expected (most likely) mark for candidates who have no knowledge whatsoever will be N/C.

Negative marking is sometimes used in order to discourage blind guessing. Usually, in a question with C options and one correct answer, 1 mark is awarded for a single correctly chosen option and -1/(C-1) for an incorrect one. In [1] this principle was extended to create the so-called "liberal" MC test format. Unknown to the author at the time, this test format had been described previously in [2], where it was referred to as a "free-choice" test.

In a liberal/free-choice test candidates are allowed to select more than one option per question, enabling them to gain a positive fractional mark for a question whenever they are able to successfully identify a subset of the options that includes the answer. Looking at it another way, they can gain a fractional mark whenever they can correctly identify one or more wrong answers for a question, even though the correct answer is unknown. In this situation we say that the candidate has *partial knowledge*, and this is rewarded explicitly in a liberal/free-choice test. It is not rewarded explicitly by number-right scoring, but it is rewarded implicitly since the candidate can now engage in *educated* guesswork rather than blind guesswork.

In [3] the author considers the effects of partial knowledge and guessing in MC tests. He differentiates between the case of lack of confidence in knowledge one actually has and the case when a correct response is guessed with greater probability of success by eliminating a distracter. By focusing on true/false tests (C=2) he avoids considering the variable probability, per question, of correct guesses and the corresponding model is simplified. In [4], the authors present empirical results based on a comparison of "number-correct" scoring and "elimination testing" scoring methods. The latter is equivalent to liberal/free-choice scoring, but candidates are asked to identify incorrect answers to a question rather than correct ones. They characterise an examinee's knowledge for a given question as one of the following: full knowledge, partial knowledge, absence of knowledge, partial misinformation and full misinformation.

In this paper our aim is to compare the conventional number-right scoring method with the liberal/free-choice scoring method. The presentation is entirely theoretical and complements the large body of literature on empirical case studies. Following Burton [5], we make the following idealised assumptions about the tests: (a) that all the questions are assumed to be of equal difficulty and well-written, (b) that the test items involve random sampling of all possible items in the relevant domain of knowledge, (c) that the candidates have sufficient time to consider all of the questions, (d) that answers to questions are given correctly whenever they are known, and (e) that in the case of number-right scoring answers are guessed whenever they are not known. As far as guessing is concerned, the usual assumption made in the literature concerning number-right scoring is that whenever candidates do not know the correct answer to a question they will guess the answer with each of the C choices having equal probability. This is where we depart from previous work. We recognise that sometimes a candidate who cannot identify the correct answer will nevertheless be able to successfully identify one or more incorrect answers due to partial knowledge, and in that case the probability of picking the correct choice will be greater than 1/C.

Partial Knowledge

First of all we consider the case when a candidate is not absolutely certain of the correct answer to a question but has some partial knowledge. We assume that s/he knows that the correct answer is one of x options, where $1<x<C$. To simplify our model even further, we assume that the candidate has equal belief in each of these x options. In a conventional MC test, the candidate has to choose just one option from C options and will get 1 mark for that question with probability $1/x$ and 0 marks with probability $\frac{x-1}{x}$. The expected mark for this question is $1/x$, and the candidate is at least as likely to score zero marks for this question as s/he is to score one mark.

Now consider the same candidate sitting an equivalent liberal/free-choice test. For this question s/he will score 1 mark for the correct option and $\frac{-1}{C-1}$ for each of the other x-1 incorrect options. The total score for this question is then $1-(x-1)\times\frac{1}{C-1}$, which equals $\frac{(C-1)-(x-1)}{C-1}$, or $\frac{C-x}{C-1}$ marks.

Comparing these marks,

$$\frac{C-x}{C-1} \geq \frac{1}{x} \qquad (1)$$

if and only if

$$x^2 - Cx + (C-1) \leq 0.$$

The roots of the quadratic equation $x^2 - Cx + (C-1) = 0$ are $x = 1$ and $x = C-1$. It can be verified easily that for values of x between 1 and $C-1$ there is strict inequality in (1). From this we can make the following observations:

- If a candidate can eliminate only one of the available options for a question ($x = C-1$), then in the liberal/free-choice test s/he will score – with probability one – the average score that s/he would have obtained by having to guess one of the $C-1$ options in a conventional MC test.

- For any other value of x in the range $1 < x < C-1$, the score for the question in the liberal MC test will be higher than the average score obtained by having to guess one of the x options in a conventional MC test.

Of course the level of partial knowledge will vary from question to question and it is difficult to model this variability for the entire test. If we assume that the level is the same for *all* of the N-K questions for which the candidate has partial knowledge (i.e. that the correct answer is one of x options), then there is a binomial probability distribution on the score that the candidate achieves in the case of the conventional MC test. Since answers to K of the questions are completely known, a score of S may be achieved, where $K \leq S \leq N$, by correctly guessing the answers to a further S-K questions. These S-K questions are chosen from the N-K questions for which the candidate has partial knowledge. Each of these S-K questions will be correctly guessed with probability $\frac{1}{x}$ while the remaining $(N\text{-}K) - (S\text{-}K)$ (or N-S) will be incorrectly guessed with probability $(1-\frac{1}{x})$. We recall that $\binom{n}{r}$ denotes the number of ways of choosing r items from n items. Then P_S, the probability that the candidate scores S marks, is given by the binomial probability distribution:

$$P_S = \binom{N-K}{S-K}\left(\frac{1}{x}\right)^{S-K}\left(\frac{x-1}{x}\right)^{N-S}. \qquad (2)$$

The mean score for the test is

$$K + \frac{N-K}{x} \qquad (3)$$

and the standard deviation is $\frac{1}{x}\sqrt{(N-K)(x-1)}$.

However, with the liberal/free-choice scoring method the total score will be

$$K + (N-K)\frac{C-x}{C-1} \qquad (4)$$

with probability one.

Comparing (3) and (4) we see that if C=3 and x=2 then a candidate with partial knowledge is not likely to achieve higher marks in a liberal/free-choice test than in a number-right test, but if $C>3$ then they would. We feel that this is a persuasive argument in favour of liberal/free-choice multiple-choice tests, but this is a moot point. The absence of a probability on the score obtained with a liberal test is undoubtedly an advantage and applies even if $C=3$. Given a particular test and a given level of partial knowledge specified by x for each question that is not completely known, although there is a probability that a candidate can score more than $K + (N - K)\frac{C-x}{C-1}$ in a conventional MC marking scheme, it is more likely that they will score less than this; whereas *this will be the score* – with probability one – in a liberal/free-choice marking scheme.

Absence of Knowledge

It is obvious that in a conventional MC test, a candidate who knows nothing can score some marks by lucky guessing. Putting $K=0$ and $x=C$ in (2) then P_s, the probability that the candidate scores S marks, is given by

$$P_s = \binom{N}{S}\left(\frac{1}{C}\right)^S\left(1-\frac{1}{C}\right)^{N-S} \qquad (5)$$

and the mean score is N/C. It is difficult to see how a candidate with no knowledge can turn the marking scheme of a liberal/free-choice MC test to his/her advantage. Under this scoring system, the candidates should be advised *not* to guess since they are as likely to lose marks as to gain them.

There are a total of 2^C ways in which s/he can choose a subset of options to a question and we will assume that all are equally likely to be chosen in the case of blind guessing, even though we realise that choosing all options is equivalent to choosing none. We will assume that each of the C options is checked with probability ½. The expected gain for checking the single correct option is ½ and the expected loss for checking each incorrect option is $-\frac{1}{2(C-1)}$. Since there are C-1 incorrect options, the average mark per question is zero and hence the average mark for the test must be zero, although of course very large positive and very large negative numbers are possible. Just how an examiner deals with an overall negative mark is another matter. Suffice it to say at this point that a candidate who blindly guesses at each question is far more likely to score zero (or less) than s/he would if sitting the same test with a conventional marking scheme. We conclude that the liberal/free-choice MC test does not reward absence of knowledge.

Misinformation

In conventional MC tests wrong answers are not punished. Thus a candidate who correctly knew the answers to K of the N questions but also falsely believed that s/he knew the answers to the remaining N-K questions would score exactly K marks out of a possible N marks. Now consider the liberal/free-choice test. We differentiate between partial misinformation and full misinformation as described in [4]. In the former case, the candidate will select y of the C-1 incorrect options and $1 \le y < C-1$; in the latter case the candidate selects all the wrong options and $y = C-1$. In some subject areas, such as medicine, it may not be desirable to tolerate such confidence in one's erroneous beliefs. In the liberal/free-choice test scenario, each of the N-K incorrectly chosen options would score $\frac{-1}{C-1}$ and so the score for the question would be $\frac{-y}{C-1}$. It can be seen that if $y=1$, the candidate is showing confidence in a single erroneous answer, and s/he will be penalised less for this question than another candidate who has full misinformation. It may be argued that this situation should be reversed. Nevertheless, in line with the degree of misinformation as described above, the candidate is punished commensurately.

A candidate who is completely confident of his/her incorrect answers will score for the test as a whole

$$K - \frac{N-K}{C-1}. \qquad (6)$$

We see that the greater the value of C, the smaller the punishment. Observe too that expression (6) is greater than or equal to zero if and only if $KC \ge N$. Thus if $K/N < 1/C$ the candidate will score zero.

Conclusions

In reality it is likely that many candidates will have partial knowledge with respect to some questions and either no knowledge or "false knowledge" with respect to other questions. Our objective in this paper has been to compare liberal/free-choice tests with conventional (number-right) tests in each of these scenarios. We conclude that liberal/free-choice tests are more generous than conventional tests in the case of partial knowledge. A candidate who can eliminate only one of the incorrect options will score (with probability one) the average score that would have been obtained by guessing one of the C-1 remaining choices with number-right scoring. Whenever a candidate can successfully eliminate two or more incorrect options then s/he is likely to achieve a higher score in a liberal/free-choice test than with number-right scoring. Given that most MC tests have C = 4 or 5, the benefits to the candidate are obvious.

Furthermore, liberal/free-choice tests do not encourage guessing. Whenever a candidate has no knowledge with respect to an individual question then s/he is as likely to lose credit as to gain credit by random guessing. This should be explained to the candidates before the test is taken, so that they are not tempted to gamble. This is of obvious benefit to the tester because it means that the resulting scores will be a more reliable indicator of what the candidates know, since they will be affected less by guesswork.

Also, any candidate who has some belief in their false knowledge will be punished. The extent to which they are punished increases with the amount of "misinformation" per question. Some of the research in MC scoring has been in the field of testing medical students. In this subject area (if not all subject areas) it is undoubtedly true that misinformation is worse than an absence of information, and it seems right that it should be penalised. Use of liberal/free-choice tests may be particularly appropriate when misinformation has life-threatening consequences and guessing is therefore to be discouraged.

References

[1] Bush, M. (2001) A multiple choice test that rewards partial knowledge. *Journal of Further and Higher Education*, 25(2), 157 -163

[2] Dressel, P. L. & Schmid, P. (1953) Some modifications of the multiple-choice item, *Educational and Psychological Measurement*, 13, 574-595.

[3] Burton, R. F. (2002) Misinformation, partial knowledge and guessing in true/false tests, *Medical Education*, 36(9), 805-811.

[4] Bradbard, D. A., Parker D. F. & Stone, G. L. (2004) An Alternate Multiple-Choice Scoring Procedure in a Macroeconomics Course. *Decision Sciences Journal of Innovative Education*, 2(1), 11 - 26

[5] Burton, R. F. (2001) Quantifying the effects of chance in multiple choice and true/false tests: item selection and guessing of answers. *Assessment & Evaluation in Higher Education,* vol 26(1) 41-50.

Citation

Jennings, S. & M. Bush (2006). A Comparison of Conventional and Liberal (Free-Choice) Multiple-Choice Tests. *Practical Assessment Research & Evaluation*, 11(8). Available online: http://pareonline.net/getvn.asp?v=11&n=8

Authors

Dr Sylvia Jennings and Dr Martin Bush are both Senior Lecturers within the Faculty of Business, Computing & Information Management at London South Bank University. They can be reached at

Faculty of BCIM
London South Bank University
103 Borough Road
London SE1 0AA