

# Penalties in multiple-choice and true-false questions

J. H. KOESLAG, C. W. MELZER, S. R. SCHACH

## Summary

In multiple-choice and true-false (T-F) questions a student expresses his ignorance by leaving questions unanswered, guessing randomly, answering according to some predetermined formula, or, in T-F tests, by marking all those which are unknown as 'true' or all as 'false'. Each of these no-knowledge strategies should, on average, yield the same score.

It is shown that partial knowledge is awarded part-marks in multiple-choice question and T-F tests, since the chance of guessing correctly is greater when students possess some knowledge than when they guess randomly. There is a constant relationship between the score obtained by the informed guesser and the minimum, maximum and random-guessing scores, this relationship being independent of the scoring system.

Modifications of the scoring system (i.e. of the magnitude of the penalty for incorrect responses) affect only the reward for unanswered questions: in the absence of a penalty the reward for unanswered questions equals the minimum score — a score which is unattainable by random guessing; with a large penalty the reward for unanswered questions approaches the maximum score obtainable.

*S Afr Med J* 1983; **63**: 20-22.

---

## Departments of Physiology and Computer Science, University of Cape Town

J. H. KOESLAG, M.B. CH.B., PH.D.

C. W. MELZER, B.SC. HONS, M.B.A.

S. R. SCHACH, M.SC., PH.D.

---

There is disagreement about whether or not marks should be deducted for incorrect responses in true-false (T-F) and multiple-choice question (MCQ) examinations. Opinions vary from those who consider the deduction of marks unnecessary or unwise<sup>1,2</sup> to those who advocate 'severe penalties'.<sup>3,4</sup> The controversy arises over the guessing factor present in T-F and MCQ examinations.

The purpose of this article is to examine the effects and anomalies created by various scoring systems.

## Total ignorance

In an MCQ or T-F examination the student who is totally ignorant of the information required to provide an answer for a particular question can: (i) not answer the question (the 'don't know' strategy); (ii) take a random guess (the 'monkey' strategy); (iii) answer according to some predetermined formula or pattern (e.g. always marking the (b) alternative as correct in 1-from-5 questions); or (iv) always mark it as true, or always as false (in T-F examinations).

Unlike the situation in an oral or in an essay examination, total ignorance is not reflected by an incorrect answer. In fact, it requires as much knowledge to get all the questions in a T-F test wrong as it does to get them all correct! In a 1-from-5 MCQ there is a probability of only  $0,2 \times 10^{-9}$  that a totally ignorant student will consistently choose one of the incorrect alternatives from each of 100 consecutive questions.

If correct and incorrect statements are sufficiently randomized ploys (ii), (iii) and (iv) will normally not be different, and the score obtained for each of them will equal the 'monkey score' (the statistically calculated score obtained by random guessing).<sup>5</sup> In certain cases, such as with the 'Middlesex Scoring Scheme',<sup>1</sup> however, true and false statements are scored differently, and an 'all-true' strategy, an 'all-false' strategy, or a 'random guessing' strategy may give rise to scores which are significantly different from each other.<sup>5</sup>

However, consistency demands that each of these no-knowledge strategies be awarded the same score (preferably zero,

to conform with convention). This is usually achieved by imposing a 'penalty' or 'countermark' for incorrect responses. The magnitude of this penalty is dependent upon the number of choices in each question, and is calculated from the formula of Guilford and Fruchter:<sup>6</sup>

$$\text{penalty} = \frac{1}{\text{number of choices} - 1}$$

Thus in a 1-from-5 MCQ the penalty would be 1/4. In a T-F test it is 1.

### Partial knowledge

The 'partial knowledge' situation is different from the no-knowledge situation and does not give the same score as the monkey strategy. Harden *et al.*<sup>1</sup> showed that when real students (as opposed to hypothetical 'monkeys') guessed in a T-F examination they gained marks even if they were penalized for incorrect answers. In other words, their chances of guessing correctly were significantly better than would have been the case if their guessing had been completely at random. This better-than-even chance of selecting the correct answer in a T-F test is a reflection of their partial knowledge, and is awarded part-marks.

In an investigation carried out by Fredman<sup>7</sup> students were asked to indicate whether they were 'absolutely sure', 'fairly sure' or 'guessing' in each question of the 1-from-5 MCQ type. A system of weighting the rewards and penalties according to the students' confidence-coding was devised to ensure truthfulness in the selection of the 'absolutely sure', 'fairly sure' and 'guessing' options. It was found that the students in the top quintile of the class achieved an average of about 44% correct answers in Fredman's examinations when they claimed that they were 'guessing', while the students in the bottom quintile got 32% of answers correct when they were 'guessing'. The brightest students were therefore making choices which were only slightly poorer than 1:2 when they were guessing.

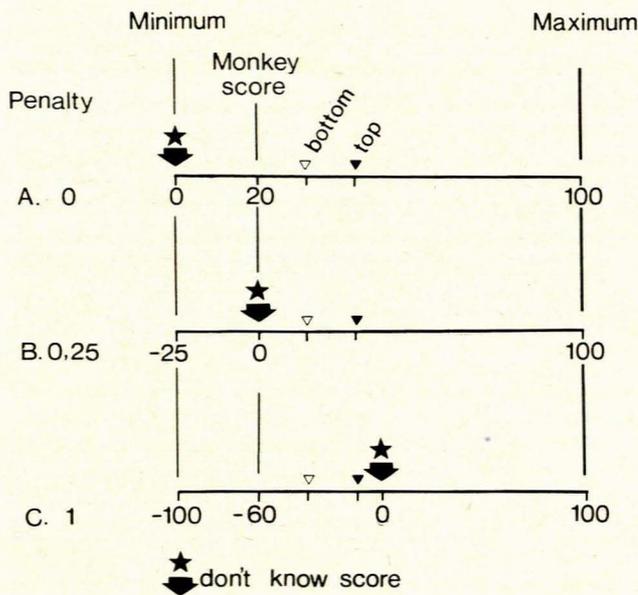


Fig. 1. Total ranges of scores in a 1-from-5 MCQ test when different penalties (0, 1/4, 1) are awarded for incorrect responses. The mean score obtained by random guessing is designated 'monkey' and is seen to occupy the same position on each line, irrespective of the magnitude of the penalty. The open arrows labelled 'bottom' and the closed arrows labelled 'top' are the scores obtained by Fredman's bottom and top quintiles of the class respectively when these students indicated that they were guessing.<sup>7</sup> The points labelled 'don't know' are the scores awarded for blank answer sheets in each case.

The poorer students also did better than statistical 'monkeys', having narrowed the choice down to approximately 3 of the 5 alternatives.

In an MCQ or T-F test 'partial knowledge' is therefore awarded part-marks, the magnitude of which is dependent on the degree of partial knowledge of the student only, and is uninfluenced by changes in the marking scheme (Fig. 1).

### The penalty

The adoption of a penalty when an MCQ or T-F test is scored has the effect of expanding the total range of the marks without altering the position of the 'monkey score' relative to the maximum and minimum scores obtainable. In a T-F test the 'monkey score' will always lie exactly midway between the minimum and maximum scores obtainable; in a 1-from-5 MCQ it lies 20% of the way between the minimum and maximum scores obtainable.

When students answer all of the questions their scores relative to one another and relative to the minimum, maximum and 'monkey' scores are also unaffected by the penalty, whatever its magnitude. Under these circumstances the penalty is irrelevant.

In many examinations, however, students are given the option or may even be encouraged to leave questions unanswered.<sup>5,8</sup> Such unanswered questions are, in nearly every case, awarded a score of zero, irrespective of the magnitude of the penalty imposed for incorrect responses. This introduces an anomaly, since zero (the 'don't know' score) may coincide with the minimum obtainable score when the penalty is omitted; it may equal the 'monkey score' if the Guilford and Fruchter<sup>6</sup> penalty is used, or it may approach the maximum score obtainable (full marks) if the penalty is very large. (Consider a scoring system which awards +1 for a correct response, -100 for an incorrect response, and zero for a 'don't know' response.)

These relationships are illustrated in Figs 1 and 2. The total range of possible scores, using different penalties in T-F (Fig. 2)

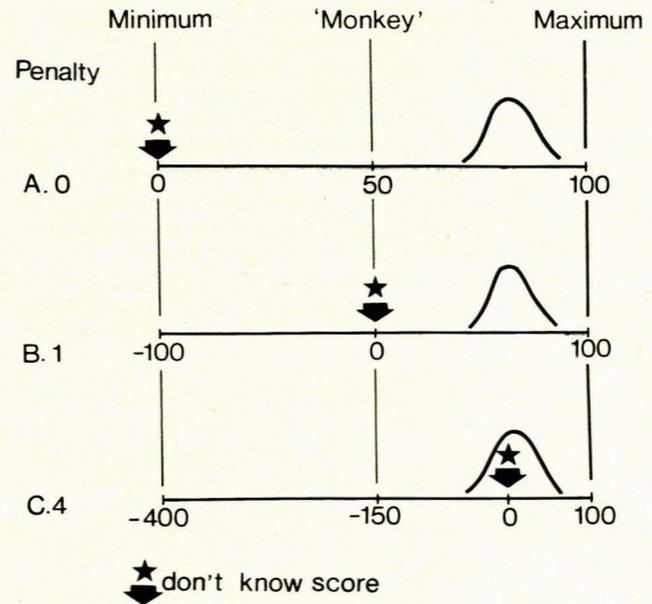


Fig. 2. Total ranges of scores in a T-F test when different penalties (0, 1, 4) are awarded for incorrect responses. The points labelled 'monkey' and 'don't know' are as in Fig. 1. The bell-shaped curve on each line represents the histogram of a set of typical scores obtained in such an examination when no student leaves questions unanswered. The position and shape of the histogram remains unaltered using different scoring schemes. Notice, however, that with a penalty of 4 marks for incorrect responses a blank answer sheet is worth almost as much as the mean mark for the class.

and 1-from-5 MCQ (Fig. 1) tests, each consisting of 100 questions, are depicted as lines of equal length. The 'monkey score' and the 'partial knowledge scores' are indicated on each line for each of the scoring systems used. For any given type of examination (T-F or MCQ) the 'monkey scores' will be seen to occupy identical positions on each line, irrespective of the magnitude of the penalty, as do the 'partial knowledge' scores of the students who answer all of the questions. Although each of these scores is given a different numerical value, depending on the size of the penalty, its percentage score remains constant and the results of the examination remain unaltered.

## Applications

These considerations apply to every conceivable system that can be devised to score student responses in all objectively evaluated tests. For no matter what format is used, what advice is given, or what scoring system is applied, in the final analysis the student sitting the examination will always be confronted with a choice of 1-from- $x$  (where  $x$  is an integer  $\geq 2$ ) for his answer. Or he may decide not to answer that question. Once this statistic (the value of  $x$ ) has been established, the maximum, minimum, no-answer and monkey scores can be calculated from the reward-and-penalty system applicable to that question or group of questions.

Questions consisting of a stem with  $y$  alternatives, of which any number may be correct or incorrect, are, in effect,  $y$ -number of separate T-F questions. In these questions the student is confronted with a choice of 1-from-2 (true or false) in each of the  $y$  alternatives.<sup>5</sup> The statistics will therefore not differ from the straightforward so-called 'independent T-F' type of test in which each question is essentially unrelated to those that precede or follow it.

Confidence-testing schemes,<sup>7</sup> or variations in the reward-and-penalty system which depend on the degree of difficulty of the question,<sup>5</sup> are also subject to the mathematical constraints described in this review. The fact that questions in these tests carry different rewards and penalties, depending on how sure or otherwise the student is of his answer, or on how difficult or otherwise the examiner considers the question to be, does not alter the fact that exact maximum, minimum, no-answer and monkey scores can be calculated for each question under the different scoring schemes. In the one case the student can select which scoring system he wishes to have applied in a given question;<sup>7</sup> in the other case the examiner makes the choice.<sup>5,7</sup> But once the choice has been made the usual arithmetic applicable to that reward-and-penalty system will rigidly apply.<sup>5</sup>

The student is advised to be aware of this, and to use it to his best advantage. We have previously described a scoring system which makes it possible for a student who has answered all the

easy questions in the examination correctly to score 3 times as many marks for all the remaining 'difficult' questions by the simple expedient of stating that they are all true!<sup>5</sup>

Of course, these statistics are completely confounded, and so are the results of the examination, if bonus marks are arbitrarily awarded to certain students for certain questions on the basis, for instance, that no group of 5 T-F statements may carry a negative score.<sup>5</sup> In such a 'lenient marking' scheme no question comprising a stem with 5 alternatives (each of which can be true or false) can yield a negative score, however badly the student fared in that group of statements. Should the reward-and-penalty system have given a student a score of -12 for those statements, he is automatically given +12 bonus marks to make his final score zero. The brighter student who scores +1 for that group of statements is not given such a bonus, and remains at +1. The result is that the inflated final scores of the below-average students are concertinaed into the main body of marks, making it difficult (if not impossible) to distinguish the below-average from the average student.<sup>5</sup>

## Conclusion

When every student answers every question in an examination, no alteration of the scoring system (except alterations in the weighting of different questions) changes their rank order or their scores relative to the minimum, maximum and 'monkey' scores.

However, when students leave questions unanswered, it is important to ensure that the score awarded for such 'don't know' responses is the same as the score obtained by the use of any of the other no-knowledge ploys. This is achieved by the imposition of a 'fair penalty' according to the Guilford and Fruchter formula.<sup>6</sup> Such a penalty does not affect the part-mark score obtained by the students who make informed guesses.

## REFERENCES

1. Harden R McG, Brown RA, Biran LA, Dallas Ross WP, Wakeford RE. Multiple choice questions: to guess or not to guess. *Med Educ* 1976; **10**: 27-32.
2. Whitby LG. Marking systems for multiple choice examinations. *Med Educ* 1977; **11**: 216-220.
3. Lennox B. Marking multiple choice examinations. *Br J Med Educ* 1967; **1**: 203-211.
4. Lennox B. Multiple choice. *Br J Med Educ* 1967; **1**: 304-344.
5. Melzer CW, Schach SR, Koeslag JH. Misconceptions and miscarriages in multiple choice questions. *S Afr Med J* 1976; **50**: 583-587.
6. Guilford JP, Fruchter B. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1973: 442.
7. Fredman M. An examination of the objective evaluation of students achievement in anatomy, with an enquiry into the results of cycling marking programs and confidence weighting of responses (Ph.D. thesis). Cape Town: University of Cape Town, 1978: 92-135.
8. Buckley-Sharp MD, Harris FTC. Scoring of multiple choice questions. *Br J Med Educ* 1971; **5**: 279-288.