Rajeeva L Karandikar
Indian Statistical Institute
7, SJS Sansanwal Marg
New Delhi 110 016, India.
Email: rlk@isid.ac.in

Multiple-Choice Tests, Negative Marks and an Alternative

I analyse various schemes of negative marking for tests consisting of multiple-choice questions and propose a scheme that reduces the impact of random guessing. I also propose an alternate style of multiple-choice questions, where each question may have several correct answers and the candidate is required to tick all correct answers in order to get credit.

Tests consisting of multiple-choice questions are very common these days. They are considered by many to be *objective* as opposed to traditional tests which are considered to be *subjective*. A basic difficulty with multiple-choice questions is that when the answer is wrong, we are sure that the candidate does not know the answer while if the answer is correct, the possibility remains that the candidate has guessed the answer without really knowing it. It is for this reason that negative marks are thought of. Some people still do not agree with the philosophy of *negative marks* for an incorrect answer. However, it is largely accepted (and most believe) that this takes care of the problem of someone getting ahead on the basis of random guessing.

But once we come to the question of what should be the negative marking scheme, there seems to be a lot of confusion. In most examinations where objective tests are used, the candidates are not told of the negative marking scheme. One argument goes as follows: *"How does it matter as to what scheme is used, it is the same for all!"*

In this article, I argue that it does matter as to what scheme is used and even if the *theoretically correct* scheme is used, candidates can get advantage using randomised guessing. I will propose an alternative negative marking

Keywords
Multiple-choice test, negative marking.

scheme, which improves matters, somewhat. Mul- tiple-choice tests are unavoidable in situations where large numbers of candidates appear and we need computerized checking of answer scripts to have a quick result. For dealing with this situation, I will also propose an alternative style of question – answers that reduce the influence of randomised guessing even further.

To make our discussion concrete, let us consider a multiple-choice test with 100 questions. Each question has 4 alternatives one of which is the correct answer. Suppose this is a screening test with 45 as the cut-off, with each correct answer getting one mark. Everyone scoring above 45 qualifies for the next round, which may be an interview or a traditional short answer test. While some examinations in this scenario use $-\frac{1}{4}$ as the score for a wrong answer, some use the *theoretically correct* $-\frac{1}{3}$ as the score for a wrong answer (this scheme has the property that the expected score of a candidate who answers all questions by randomly choosing an alternative is zero – in this sense, it is theoretically correct).

We will call scheme I the marking scheme with no negative marking, scheme II as the scheme with $-\frac{1}{4}$ as the score for a wrong answer and scheme III as the scheme with $-\frac{1}{3}$ as the score for a wrong answer. In each of these, the correct answer gets a score of 1.

Let us consider a candidate, named **A** in the sequel, who knows answers to 40 questions while he randomly guesses the answers to the remaining 60. In this case, the expected score of **A** is 55 under scheme I; 43.75 in scheme II and 40 in scheme III. A little reflection will convince us that it is not the expected score but the *probability* of **A** scoring 45 or more that is relevant – if there are a large number of candidates appearing in an examination and suppose there are 1000 candidates like **A**, what is relevant is not the average score of these 1000 candidates (which will be close to 55, 43.75 and

40 under the three schemes), but *how many* out of these will have a score above 45. In this situation, the number of correct answers X_A (out of 60 randomly guessed) by **A** has a binomial distribution with $n = 60$ and $p = 0.25$ and so the probability of net score being over 45 can be computed: it is $P(X_A \geq 5)$ under scheme I; $P(X_A \geq 16)$ under scheme II and $P(X_A \geq 19)$ under scheme III. These probabilities are 1, 0.43 and 0.15.

This shows that while we must have negative marks, under the scheme II, a candidate can make up a gap of 5 marks with about 43% probability. Even with scheme III, the probability is nearly 15%. So by guessing, a gap of 5 marks has been made up with high probability. Let us see the risk involved in randomised guessing for a candidate **B** who has 5 extra marks, i.e. knows answers to 50 questions and answers the remaining by guessing. In this case the probabilities are 1, 0.99 and 0.91 respectively. So under scheme II there is 43% chance of gaining 5 marks and about 1% chance of losing 5 marks while under scheme III, there is 15% chance of gaining 5 marks and about 9% chance of losing 5 marks.

Suppose there is a candidate **Z** who answers 44 questions correctly and does not guess any answer. In this case, the chances that **A** will score more than **Z** does depend upon the scheme – under scheme I it is 1 while under II it is 0.43 and under scheme III it is 0.15. So the negative marking scheme does have an impact upon the ordering of the candidates.

People who have participated in setting questions for such tests will agree that finding 3 credible alternatives is often rather difficult. In many questions, one of the alternatives is such that it can be ruled out with little or no knowledge and hence we should not give credit to a candidate who has ruled out one of the answers in many questions. So let us consider a candidate **C** who answers 40 questions based on his *knowledge*. In 50 questions he rules out correctly one alternative and chooses one out of three alternatives randomly and remaining 10 questions he has no clue and chooses an answer randomly out of 4 alternatives. As we have argued above, we would not like to select him for next round, as we do not wish to give credit to his having ruled out one alternative in 50 questions. In this case, the number of correct answers X_C out of the ones **C** has guessed no longer has a binomial distribution but is a sum of two binomial distributions with parameters $n = 50$, $p = 0.33333$ and $n = 10$, $p = 0.25$ respectively. The distribution of the sum can be worked out but it is much easier to compute the probability that X_C is at least 45 via simulation (see [1]). They are 1 under scheme I, 0.85 under scheme II and 0.57 under scheme III. Let us also examine a candidate **D**, who knows answers to 50 questions, in 40 questions he rules out correctly one alternative and chooses one out of three alternatives randomly and remaining 10 questions he has no clue and chooses an answer randomly out of 4 alternatives. For **D**, the probability that X_D exceeds 45 are 1,1 and 0.99 under the three schemes respectively.

Thus, if we think that in most questions, one option can be ruled out without much knowledge of the subject matter, or to put it differently, we do not wish to give partial credit to a candidate for ruling out one alternative out of 4; then we should not use scheme III. For a candidate can have a strategy of randomly guessing out of 3 alternatives in all questions that the candidate does not know the correct answer but can rule out one alternative. With this strategy, the candidate can make up a shortfall of 5 marks with probability 0.57 under scheme III but the probability of losing 5 marks is only 0.01.

We can think of scheme III as follows: A random guess out of 4 does not get any credit (on the average) while ruling out one alternative correctly and guessing out of the remaining three gets a credit of $\frac{1}{9}$ on the average

($\frac{1}{3}(1 - \frac{1}{3} - \frac{1}{3})$) while ruling out 2 options correctly and guessing out of the remaining two gets a credit of $\frac{1}{2}(1 - \frac{1}{3}) = \frac{1}{3}$ on the average.

In this light, consistent with the view that ruling out one alternative should not get any credit, while ruling out two alternatives should get some partial credit, we can consider a scheme IV where an incorrect answer gets a score $-\frac{1}{2}$. Under this scheme, a student will be penalised if he randomly guesses out of four alternatives, while gets a score of $\frac{1}{4}$ on the average for guessing out of two alternatives after ruling out (correctly) two alternatives.

Table 1 gives the probability of scoring 45 or more under the three schemes with $-\frac{1}{2}$,$-\frac{1}{3}$ and $-\frac{1}{4}$ marks for a wrong answer. Here, $(a : b : c : d)$ represents a student who has answered correctly a questions knowingly, has guessed out of two alternatives in b questions, guessed out of three alternatives in c questions and guessed out of four alternatives in d questions.

Table 1.

$(a : b : c : d)$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{4}$
(35:0:65:0)	0.038	0.31	0.61
(35:50:10:0)	0.68	0.94	0.99
(35:60:5:0)	0.78	0.98	1
(38:0:0:62)	0.0026	0.074	0.27
(38:0:60:2)	0.09	0.49	0.79
(38:40:4:0)	0.71	0.93	0.98
(38:48:4:0)	0.86	0.97	1
(38:48:4:0)	0.86	0.97	1
(38:58:4:0)	0.89	0.99	1
(40:0:0:60)	0.0074	0.15	0.43
(40:0:50:10)	0.12	0.57	0.85
(40:20:10:0)	0.47	0.75	0.92
(40:25:10:0)	0.68	0.87	0.97
(40:30:0:0)	0.71	0.9	0.98
(50:0:0:50)	0.36	0.91	0.99
(50:0:40:10)	0.76	0.99	1

We strongly recommend the scheme where a wrong answer gets $-\frac{1}{2}$. The table given above shows that under this scheme, a candidate can hope to gain significantly only when he/she can rule out two alternatives on a given question and then guess randomly out of the remaining two. Here, we cannot eliminate the effect of random guessing, but the effect is minimized.

This brings us to another question. Consider a situation where multiple-choice test is being used, and the test is conducted say twice or thrice a year and candidates can appear again and again. There are many such situations. So in about a year, a candidate may have several chances, say 3. This drastically pushes up the chances of a candidate getting selected in one of the tests. So let us consider a test, which is to be used as a final selection test (and so the cut-off is high, say 65) and a candidate can appear 3 times in the exam. The table below gives probability of selection (scoring 65 or more) in up to 3 attempts. As in the earlier case, $(a : b : c : d)$ represents a student who has answered correctly a questions knowingly, has guessed out of two alternatives in b questions, guessed out of three alternatives in c questions and guessed out of four alternatives in d questions (in each of his attempts). *Table* 2 gives the probability of a candidate scoring 65 or more in at least one of the three attempts.

We can see that with $-\frac{1}{3}$ as marks for an incorrect answer, shortfall of 10 marks can be made-up with a high probability, and if a candidate can rule out two alternatives in many questions, a shortfall of 10 can be made up with a high probability even with $-\frac{1}{2}$ as marks for an incorrect answer.

Let us examine an alternative. Recall that the main reason in favour of using multiple-choice questions is that the evaluation can be done by scanning the answer sheet and then computerized marking. This enables the

Table 2.

$(a:b:c:d)$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{4}$
(50:10:40:0)	0.026	0.18	0.43
(50:20:30:0)	0.086	0.39	0.7
(50:30:20:0)	0.22	0.65	0.89
(50:40:10:0)	0.45	0.86	0.97
(50:50:0:0)	0.71	0.96	1
(55:0:40:5)	0.046	0.29	0.61
(55:5:40:0)	0.11	0.49	0.8
(55:10:35:0)	0.19	0.63	0.89
(55:15:30:0)	0.29	0.76	0.94
(55:20:25:0)	0.42	0.86	0.98
(55:25:20:0)	0.56	0.92	0.99
(55:30:15:0)	0.69	0.97	1
(60:0:40:0)	0.37	0.85	0.98
(60:5:35:0)	0.52	0.93	0.99
(60:10:30:0)	0.67	0.97	1
(60:15:25:0)	0.79	0.99	1

agency conducting examination to process large number of answer scripts in short time. One alternative to reduce the effect of randomised guessing is to increase the number of alternatives for each question, but as argued above, finding credible alternatives is not easy. Another option is to have questions for which several options may be correct and the candidates are required to mark all correct alternatives. So while we may still have 4 alternatives, the effective alternatives rises to 15 (we should keep at least one correct answer so that we can differentiate between attempted and unanswered). Then we can do away with negative marking. Let me explain this by an example. The example is a question in mathematics. Subsets of real numbers can have many properties such as being *open* and/or *connected* and/or *bounded* and/or *dense*. Consider the question:

Tick all correct answers: The set of real numbers $\{x \in \mathbb{R} : (\sin(x))^2 < 1\}$ is

(a) Open; (b) Connected; c) Bounded; (d) Dense.

Here the answer is (a) and (d). The point is if a candidate is clueless and is going to tick at random, he has to consider all possibilities (15 of them) and chance of ticking the correct one is only $\frac{1}{15}$.

Consider another question: a certain numerical question in physics or chemistry, whose answer is say 3.23. The question could be: Tick all correct answers:

(a) between 1.1 and 2.8; (b) between 1.9 and 3.4;

c) between 2.5 and 3.76; d) between 3.1 and 3.9.

Here the candidate should tick three options (b), (c) and (d).

In a given test, the examiners should take care that there are enough questions with 1, 2, 3 as well as all 4 correct answers. This way, a guessing strategy would not give an advantage to a candidate, as effectively there are 15 options.

Here, we may also consider alternate schemes, where each option may have associated with it a number x – the marks a student gets for ticking that option (x positive for correct choices and negative for incorrect choices). The marks for correct choices should add to 1 on each question. For example, for the question given above with (a), (d) as correct answers, the marks could be $\frac{1}{2}$ for (a), $-\frac{1}{2}$ for (b), $-\frac{1}{2}$ for (c) and $\frac{1}{2}$ for (d). The questions could be such that they require some analysis or computation rather than test memory. It may be noted that the questions in this new style would typically require more time for a candidate to understand and answer. Thus, the candidates should be given more time to answer. This way, the test will be able to give better discrimination based on knowledge base of the candidates than the present style of multiple-choice tests, which often end up being a test of memory and speed.

Suggested Reading

[1] Sudhakar Kunte, Technique of statistical simulation, *Resonance*, Vol.5, No.4, 2000.