A NEW SCHEME FOR MULTIPLE CHOICE TESTS IN LOWER DIVISION MATHEMATICS

by

BRUCE R. JOHNSON

DMS–516–IR

September 1989

A NEW SCHEME FOR MULTIPLE CHOICE TESTS IN LOWER DIVISION MATHEMATICS

BRUCE R. JOHNSON

Department of Mathematics and Statistics
University of Victoria
Victoria, B.C., Canada
V8W 2Y2

With the class enrollment figures for my lower division mathematics and statistics classes rising over the hundred mark, I reluctantly made the decision to experiment with multiple choice tests. I began with a 50–minute midterm test consisting of 20 multiple choice questions each in the standard five–choice format. It was a pleasure to experience the speed and accuracy of computer grading with the added bonus of question–by–question statistical analysis of class performance. I felt reasonably comfortable about the additional emphasis on error–free work due to no partial credit awards. In our high tech age it seems that small technical errors are often harder to spot and potentially more damaging than ever before, so it probably is appropriate to put more weight on accuracy in first–year and second–year mathematics courses. Also, I found that multiple choice provided a framework for broader, but less deep, coverage of the concepts because of its suitability for many short questions rather than a few long ones. However, I found much that troubled me more than the lack of depth. Setting the test was a slow, painful, mind–straining chore. It was very tiresome trying to anticipate, for each question, the four most common incorrect numerical answers to accompany the one correct answer. I was further troubled by the thought that some students would work backwards from the answer choices using a process of elimination and that guessing as a last resort would yield at least a 20% chance of success. But what bothered me most of all was the realization that the student who arrived at an answer which was not listed would be warned automatically of a mistake, while the student who arrived at a listed wrong answer would not. To warn some and not others is unfair. I rejected the possible solution of always having one of the five choices "none of the above" because of its unacceptable behavior as the correct choice rewarding every numerical answer except the four listed wrong answers.

Now I have devised an approach which substantially reduces the shortcomings of standard multiple choice tests. This approach is especially suitable for lower–division mathematics and statistics where most of the questions have numerical answers. For each multiple choice question requiring a numerical answer, ten possible choices are given in numerically increasing order. The correct answer might be one of the ten choices, but usually is not. The students are to select the choice nearest their computed answer. In the event of their answer being equidistant from the two nearest values given, the rule is to select the larger of these two choices.

Questions are much easier to set in the new format. The choices should be spaced over a feasible range at intervals narrow enough to provide a good chance of catching student errors. From question to question the position of the correct choice should be randomly shuffled among primarily the interior eight choices. Often the smallest and largest choices make undesirable correct answers because either setup might fail to detect a wide range of wrong answers. However, it is important for the instructor to use the extreme choices once in a while to avoid being predictable. I usually set either the smallest or largest choice as correct answer for one or two of the easiest questions where virtually everyone in the class is expected to arrive at the correct answer.

With choices in numerically increasing order, the instructor's task of choosing order is trivial and the student's task of selecting the appropriate choice based on computed answer is quick and easy.

In the ten–choice nearest–answer framework, students are unable to work backwards from the answers using a process of elimination, and guessing carries only a 10% chance of success. While on the subject of guessing, I would like to register my strong opposition to any scoring scheme that attempts to discourage guessing by penalizing wrong answers more severely than omitted answers. With multiple choice questions it is impossible to separate blind guesses from honest errors, and most honest errors deserve more credit than omitted answers, not less. In my opinion the most fair compromise is to award no credit for either wrong or omitted answers.

Although this multiple choice format is a vast improvement over the standard format, I do not believe that any multiple choice test is capable of adequately handling in–depth questions. To deal with this problem I currently use tests for my lower division mathematics and statistics classes that are part multiple choice and part full answer. The multiple choice part, worth from one–half to two–thirds of the total, typically consists of several short questions giving broad coverage of the concepts; the full–answer part consists of a few more involved questions. I believe this combined format offers a good solution for large enrollments and the increasing need to emphasize technical accuracy.

With most of the easy questions appearing in the multiple choice section, it is not surprising that my students have tended to score higher on this section than on the full answer section despite the possibility for partial credit in the latter. However, wanting to obtain a fair statistical comparison, I set one combined test where I tried to create questions of comparable difficulty for both sections. This test was taken by 101 students. The sample mean scores were 53.25% and 55.11%

for the multiple choice and full–answer segments, respectively. The sample of 101 paired differences (percentage score for multiple choice segment minus percentage score for full–answer segment) yielded a sample mean difference of –1.86% with estimated standard error = 2.31%. Thus, there was no statistical evidence (two–sided p–value = 0.418) that the true mean difference was other than zero. On average, students tended to perform the same on the multiple choice segment as on the full–answer segment.

To familiarize students with the new test format, I have adopted the routine of distributing practice tests having the same format.

Below are some examples of test questions taken from recent examinations. The first two questions were taken from a Finite Mathematics test, and the last two from one in an Introduction to Probability and Statistics.

1. Find the first row, second column entry of M^{-1} where

$$M = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}.$$

(A) Does not exist	(B) –1.0	(C) –0.5	(D) 0.0	(E) 0.5
(F) 1.0	(G) 1.5	(H) 2.0	(I) 2.5	(J) 3.0

$$M^{-1} = \begin{bmatrix} 0.4 & -0.6 \\ -0.2 & 0.8 \end{bmatrix}.$$

2. A research team consists of 9 men and 6 women. A project outline is to be developed by a group of 3 men and 2 women from the team. How many such groups are possible?

(A) 100	(B) 200	(C) 300	(D) 400	(E) 500
(F) 600	(G) 900	(H) 1200	(I) 2400	(J) 4800

The number of groups is $\begin{bmatrix} 9 \\ 3 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = 1260.$

3. A certain fabric has an average of 1.5 flaws per square meter. In a 6–square–meter piece of this fabric, what is the probability that there are no more than 11 flaws?

(A) 0.0 (B) 0.1 (C) 0.2 (D) 0.3 (E) 0.4
(F) 0.5 (G) 0.6 (H) 0.7 (I) 0.8 (J) 0.9

Since $\Sigma_{k=0}^{11} \frac{9^k}{k!} e^{-9} = 0.803$, the correct choice is (I).

4. Let X_1, X_2, X_3, X_4 be a random sample from a distribution with mean $\mu = 2.4$ and standard deviation $\sigma = 3.0$. Find the standard deviation of $\bar{X}$.

(A) 1.0 (B) 1.5 (C) 2.0 (D) 2.5 (E) 3.0
(F) 3.5 (G) 4.0 (H) 4.5 (I) 5.0 (J) 5.5

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{4}} = 1.5.$$