Article from:

The Actuary

December 1980 – Volume 14, No. 10

The Actuary

The Newsletter of the Society of Actuaries

SUPPLEMENT NOVEMBER, 1980

SCORING MULTIPLE-CHOICE ACTUARIAL EXAMINATIONS

by R. Stephen Radcliffe and Terry A. Nicodemus

For a number of years, exam scores for Parts 1, 2 and 3 have been based on the number of correct answers <u>plus</u> one-fifth of the number of questions omitted (Method 1). However, until May 1979, the multiple choice scores on all other exams were based on the number of correct answers <u>minus</u> one-quarter of the number of incorrect answers (Method 2).

Starting in November 1979, exams which are 100% multiple choice are being graded uniformly using Method 1. Method 2 will continue to be used on exams that contain both multiple choice and essay questions.

Method 1 and Method 2 are equivalent in ranking the candidate's multiple choice scores. Therefore, <u>exactly</u> the same candidates will pass or fail under either method.

Equivalence of the two methods can be demonstrated as follows:

Let T = the number of questions on the exam

R_t = the number of correct answers by candidate t

W_t = the number of incorrect answers by candidate t

O_t = the number of questions omitted by candidate t

S_t^1 = the score produced by Method 1

S_t^2 = the score produced by Method 2

By definition,

$$S_t^1 = R_t + .2(O_t)$$

$$S_t^2 = R_t - .25(W_t)$$

$$T = R_t + W_t + O_t$$

For two candidates a and b, it can be shown that $S_a^1 \leqslant S_b^1$ if and only if $S_a^2 \leqslant S_b^2$; i.e., $R_a + .2(O_a) \leqslant R_b + .2(O_t)$ if and only if $R_a - .25(W_a) \leqslant R_b - .25(W_b)$.

As a matter of fact, these scores are related by a linear transformation.

Since $5(S_a^1) = 5(R_a) + O_a$,

and $4(S_a^2) = 4(R_a) - W_a$,

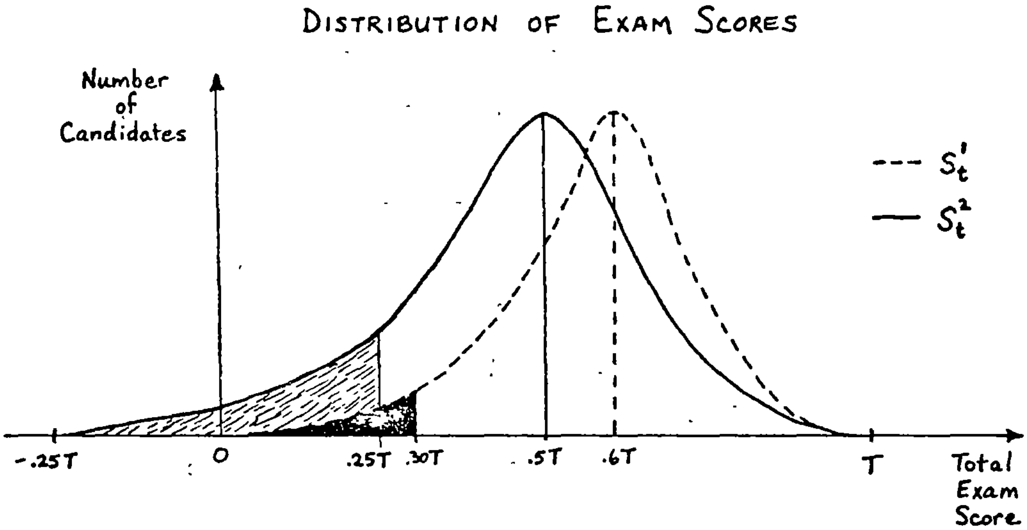
we have $S_a^1 = .8(S_a^2) + .2(T)$.

(Continued on back)

Method 1 increases the scores without changing the ranking of candidates. Furthermore, the difference in scores for the same candidate, using each of the two methods, increases as the scores get lower because $S_t^1 - S_t^2 = .2(T - S_t^2)$.

Although under each of these methods the same candidates pass or fail, the scope of the decile grades (0 to 10) will differ slightly. This is because the range of scores is altered. The range for S_t^1 is from zero to T, while the corresponding range for S_t^2 is from - .25T to T, which is a wider range of scores than under Method 1. It also follows that the median and mean scores will be lower under Method 2.

The graph below shows an example of how the distributions of scores by the two methods range from the minimum up to the maximum.



For the example, suppose the Method 2 pass mark is set at .5T (the hypothetical median in the table above). Under Method 1, the corresponding pass mark will be .6T. The solid line shows the distribution of scores under Method 2, and the dotted line the corresponding distribution of scores under Method 1. The shaded areas show that the proportion of candidates classified as ineffective (those receiving a grade of zero) is significantly smaller under Method 1 than under Method 2. Continuing this line of logic, the graph also illustrates how the change in scoring methods also changes the scope of grades other than zero.

To summarize, the critical result - whether the candidate passes or fails — has not changed. But under Method 1, fewer candidates receive the high and low grades (e.g., zero and 10), while more candidates receive median grades (e.g., 5 and 6). □