BIAS AND ERROR IN MULTIPLE-CHOICE TESTS

C. HORACE HAMILTON
NORTH CAROLINA STATE COLLEGE

A formula for estimating real scores on a multiple-choice test from a knowledge of raw scores is derived. This formula does not involve the assumption of a binomial distribution of real scores as does the Calandra formula. Other important formulas derived show: the variance of real scores in terms of the variance of raw scores and the correlation between real scores and raw scores. If the variance of real scores (or of raw scores also) is binomial, the regression of real scores on raw scores is linear; but, otherwise the regression is curvilinear. Yet the linear estimating formula is a close approximation to the curvilinear relationship. Factors affecting the regression of real scores on raw scores and the correlation coefficient are: (1) the number of choices per question; (2) the number of questions answered; (3) the ratio of the average group raw score to the variance of raw scores.

The major purpose of the multiple-choice type of objective test is to eliminate human error in grading or scoring. However, such tests are peculiarly subject to a bias and an error not present in non-choice objective tests. This situation prevails because students usually guess at answers they do not know — in spite of unrealistic instructions to the contrary. For the purpose of this paper, we shall assume that students taking multiple-choice tests will guess at all the answers they do not know. As a matter of fact, there is no good reason why they should not be encouraged to guess. Possibly there is no clear-cut line between an informed answer and a shrewd intuitive guess.

Various methods have been proposed for scoring multiple-choice tests, and certain theoretically unsound formulas for adjusting test scores are widely used. Unfortunately, very few constructive theoretical contributions have been made to this problem, and the best contributions have been almost completely ignored. Perhaps the most valuable paper on the subject is that of Alexander Calandra, published in 1941 (1). Calandra made Bayes' theorem of inverse probability the basis for scoring multiple-choice examinations. However, there is little, if any, evidence that his scoring formula has been made use of by educators and professional testers.
There is, of course, a long history of empirical research and experimentation with scoring methods. The tendency has been to ignore purely theoretical contributions. It must be granted that any and all theory should be tested and checked by empirical methods; but on the other hand, it should likewise be conceded that our empirical research will be facilitated if it takes full advantage of the advancements in theory.

It is not possible in one short paper to review all the important contributions to the problem of scoring multiple-choice tests. It is interesting to note, however, that over a period of at least 27 years there has been in vogue a scoring formula which is not based on sound mathematical theory (2, 3). This formula is also widely used by professional examination concerns. The formula is:

\[ S = R - \frac{W}{k-1}, \quad (1) \]

where

- \( S \) is the adjusted or real score,
- \( R \) is the unadjusted or raw score,
- \( W \) is the number of incorrect answers, and
- \( k \) is the number of choices per question.

Other forms of this formula are:

\[ S = n - \frac{kw}{k-1} = \frac{kr-n}{k-1}, \quad (1a) \text{ and } (1b) \]

where \( n \) is the number of questions in a test.

The above formula is apparently based upon a misunderstanding of inverse probability. In correlation terminology, it purports to be a symbolic statement of the linear regression of real scores on raw scores. As a matter of fact, it is nothing more nor less than an improper reversal of the formula for the actual linear regression of raw scores on real scores. As usually stated, it does not even identify \( S \) as an estimated score nor do we find anywhere any reference to an error of estimate.

For the most part, in this paper we shall use the symbols, terminology, and concepts of correlation and regression.

In a group test situation, there will be \( N \) students taking a test of \( n \) questions and having \( k \) choices per question. Unless otherwise indicated, we shall assume that all students will answer all questions