THE EFFECT OF DIFFICULTY AND CHANCE SUCCESS ON ITEM-TEST CORRELATION AND ON TEST RELIABILITY*

LYNNETTE B. PLUMLEE
EDUCATIONAL TESTING SERVICE

An equation is derived for predicting the effect of chance success, relative to item difficulty, on item-test correlation. The values predicted by this equation and by equations derived by Guilford and Carroll for predicting the effect of chance success on item difficulty and test reliability are compared with empirical values in an experiment which used identical test items in multiple-choice and answer-only form.

Introduction

The "multiple-choice" type of test, in which answer options are supplied, frequently has been objected to on the grounds that an examinee who does not know the answer to any of the items in a test can make a substantial score by pure guesswork and that his score therefore does not represent his true knowledge. In support of the multiple-choice test, on the other hand, it has been argued that this type of test can be made both a more effective and a more efficient measuring instrument than the "answer-only" type of test, in which no answer options are given and which thus require the examinee to supply his own answers. Since wrong answer options can be restricted to answers resulting from certain types of errors, the multiple-choice item can be directed towards testing specific types of errors. The multiple-choice test places the burden of decision regarding the correctness of an answer on the examinee rather than on the scorer, and greater consistency in scoring is thus possible in this type of test than in the "answer-only" test.

One aspect of the problem is investigated in this paper: What is

*Condensation of a dissertation presented in partial fulfillment of the requirements for the Ph.D. degree to the University of Chicago. Grateful acknowledgment is made to Professor Harold Gulliksen for his guidance as thesis advisor and to Professor L. L. Thurstone and Dr. D. W. Fiske of the University of Chicago who served as members of the thesis committee. The author is also indebted to Professor S. S. Wilks for review of the derivations and development of statistical tests used in the thesis, to Dr. L. R Tucker for technical advice, and to Dr. W. G. Mollenkopf for critical comments on the derivations and interpretations. The writer expresses appreciation to the Educational Testing Service for making available its technical facilities, and to the University of Chicago for the flexible administrative arrangement which made this thesis possible.
the theoretical effect of chance success on item-test correlation and on test reliability, and to what extent is this theoretical expectation borne out in practice?

Several previous investigations (1-13, 15) have been concerned with the prediction of the effect of chance success on item difficulty, inter-item correlation, item-test correlation, and test reliability.

In 1936, Guilford (3) presented a formula for the proportion of examinees who know the answer to an item, \( p \), as a function of the proportion of examinees who answer the item correctly in multiple-choice form, \( p \), and the number of answer options, \( n \):

\[
\phi = \frac{np}{n-1} - \frac{1}{n-1}.
\] (1)

This formula assumes that all answer options are equally attractive to the examinee who does not know the correct answer.

Carroll (1) considered the combined effect of item difficulty and chance success on the Pearsonian correlation between items or between sets of items. Using the binomial distribution of chance failure (number-not-right) scores, he arrived at the following formulas for mean, standard deviation, inter-test correlation, and inter-item correlation, respectively:

\[
E' = W \bar{E},
\] (2)

\[
\sigma_{E'} = \sqrt{W^2 \sigma_{E}^2 + RWE},
\] (3)

\[
\sqrt{Wa_{1}a_{2} \sigma_{E_{1}} \sigma_{E_{2}}} \frac{W_{a_{1}a_{2}} \sigma_{E_{1}} \sigma_{E_{2}}}{\sqrt{W^2 \sigma_{E_{1}}^2 + RWE_{1} \sigma_{E_{1}}^2 + RWE_{2} \sigma_{E_{2}}^2 + R^2 E_{1} E_{2}}},
\] and (4)

\[
r_{1,2'} = \frac{Wq_1 (1 - q_2)}{\sqrt{q_1 q_2 (1 - Wq_1) (1 - Wq_2)}},
\] (5)

where \( E \) represents the failure score, \( W \) is the probability of failure and \( R \) is the probability of success in multiple-choice form, a prime denotes a multiple-choice test, absence of a prime denotes an answer-only test, subscripts 1 and 2 denote two different items or tests, and \( q \) is the probability of failing a given item in answer-only form. (Notations used by the present author have been substituted for some of those used by Carroll.)

It is the plan of the present paper to develop the equation for