DOCUMENT RESUME

ED 137 354 TM 006 155

AUTHOR Gavin, Anne T.; Martin, Charles G.

TITLE A Procedure for Estimating the Unique Contribution of Each Component of a Composite Test: Uniqueness Analysis of Test 500. Technical Memorandum 76-8.

INSTITUTION Civil Service Commission, Washington, D.C. Personnel Measurement Research and Development Center.

PUB DATE May 76

NOTE 14p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.

DESCRIPTORS Aptitude Tests; Correlation; *Occupational Tests; Scores; *Statistical Analysis; *Test Reliability; Test Validity; True Scores; Weighted Scores

IDENTIFIERS *Composite Tests; Professional Administrative Career Exam Test 500; Test Score Variance; *Uniqueness Analysis

ABSTRACT

A procedure for estimating the degree to which a subtest uniquely contributes to total test performance is presented and discussed. Uniqueness analysis may be appropriately applied to any composite measurement instrument such as a multipart test or a multitest battery to assess the unique contribution of each component to the total test. The presentation in this report is in terms of the applicability of these procedures to Test 500 of the Professional and Administrative Career Examination. Uniqueness analyses were conducted on each series of Test 500 administered competitively in FY 75. The results for each of the five subtests, showing the degree to which subtest variance can be attributed to error, overlap, and unique sources, and indicating the potential unique validity of each subtest are presented by test series. (Author)

Personnel Research and Development Center

Technical Memora

A Procedure for Estimating the Unique Contribution of each Component of a Composite Test: Uniqueness Analysis of Test 500

United States
Civil Service Commission
Bureau of Policies and Standards

Technical Memorandum 76-8

A PROCEDURE FOR ESTIMATING THE UNIQUE CONTRIBUTION OF EACH COMPONENT OF A COMPOSITE TEST: UNIQUENESS ANALYSIS OF TEST 500

Anne T. Gavin
and
Charles G. Martin

Test Services Section
Personnel Research and Development Center
U. S. Civil Service Commission
Washington, D. C. 20415

May 1976

A PROCEDURE FOR ESTIMATING THE UNIQUE CONTRIBUTION OF EACH COMPONENT OF A COMPOSITE TEST: UNIQUENESS ANALYSIS OF TEST 500

ABSTRACT

A procedure for estimating the degree to which a subtest uniquely contributes to total test performance is presented and discussed. Uniqueness analysis may be appropriately applied to any composite measurement instrument such as a multipart test or a multitest battery to assess the unique contribution of each component to the total test. The presentation in this report is in terms of the applicability of these procedures to Test 500 of the Professional and Administrative Career Examination. Uniqueness analyses were conducted on each series of Test 500 administered competitively in FY 75. The results for each of the five subtests, showing the degree to which subtest variance can be attributed to error, overlap, and unique sources, and indicating the potential unique validity of each subtest are presented by test series.

PREFACE

The purpose of this Technical Memorandum is to present and demonstrate a procedure for determining the proportion of unique and reliable variance in a subtest that is being used in a linear composite test or in a test battery. It is designed primarily for psychologists of the U. S. Civil Service Commission for use in test development. The uniqueness analysis procedure was applied to the four series of Test 500 of the Professional and Administrative Career Examination administered in FY 75 and serves as part of the documentation of this test.

TABLE OF CONTENTS

A PROCEDURE FOR ESTIMATING THE UNIQUE CONTRIBUTION OF EACH COMPONENT OF A COMPOSITE TEST: UNIQUENESS ANALYSIS OF TEST 500

Test 500 of the Professional and Administrative Career Examination (PACE) is a written test of five abilities shown by job analytic research to be necessary for successful performance in a number of Federal occupations. The five subtest scores are differentially weighted for six different patterns of occupations. A linear composite of the weighted ability scores is used to determine an applicant's examination rating for each pattern (Wing, 1974).

Underlying the composite nature of Test 500 and the differential scoring procedure is the assumption that each of the five ability subtests makes a unique contribution to the descriptive and predictive efficiency of the total test. It is assumed that performance on each subtest relies to some degree on abilities not tapped by the other subtests; each subtest score indicates some unique aspect of an applicant's abilities. Therefore, information obtained from one subtest score does not replicate information from the other subtest scores in describing applicants' abilities and ranking them accordingly. Furthermore, since each of the subtests has been designed to measure an ability shown to be necessary for successful job performance, a subtest score which uniquely contributes to total test performance also adds to the validity of the total test against criteria of job performance.

This Technical Memorandum provides test users and psychologists with a comprehensive procedure for assessing the unique contribution of components of a composite measurement instrument. Basic to the understanding and derivation of this procedure is the concept of variance in test scores; i.e., the sources of test variance and methods for determining how much of the variance in a composite test is attributable to each component source should be understood. These concepts are discussed in detail and statistical methods for their quantification are given.

Techniques for assessing the unique contribution of each subtest of a composite test were applied to each of the four series of Test 500 administered competitively in FY 75 (Series 110, 120, 130, and 140). The determination of the variance components for each subtest are presented and discussed as well as the results of the uniqueness analyses of each subtest of each series.

COMPONENTS OF VARIANCE IN TEST SCORES

When tests are reliable, variance in scores results for the most part from individual differences among the competitors taking the test (Guilford & Fruchter, 1973, p. 68). Not all the variation in scores is due to such individual differences however, because a single test score is rarely a perfect indication of a person's true score on the variable in question. In the classical test theory model (Lord & Novick, 1968, p. 31), this discrepancy between true score and observed score is attributed to error arising from imperfect measuring instruments (errors of measurement).

It is important for test users to know how much of the variance in observed test scores is attributable to variance in individuals' true scores. For test scores to differentiate among competitors, it is necessary that there be variability in their true scores. The true score variance is reflected in the observed score variance. If the observed score variance arises from error rather than from real differences among the competitors however, their relative placement along the range of scores is based on chance and gives no meaningful information about their abilities.

Procedures for estimating the reliability of a test deal with the issue of true and error variance in test scores. Although there are a number of statistical procedures for estimating reliability, each arising from a different conception of the sources and types of errors of measurement and therefore leading to different interpretations, reliability can be defined as the proportion of observed score variance of a set of measurements that is true variance (Guilford & Fruchter, 1973, p. 397). Thus, the reliability coefficient of a test indicates the proportion of the variance of observed scores that is attributable to true differences among competitors.

The concept of test score variance becomes more complex when dealing with a composite measurement instrument. In this case, the total test variance is a function of the variances and covariances of the subtests. Nonzero covariances among the subtests indicate that these subtests share a true score variance component. (See Martin, 1975 for a more complete treatment of this topic.) The total variance of each subtest can be considered as the sum of variance from three sources: (1) that common to more than one of the subtests; (2) that unique to the subtest; and (3) that due to errors of measurement. The third component has already been discussed. Only the reliable subtest variance can be meaningfully partitioned into common and unique components.

Partitioning Variance

Correlation analysis serves to partition the variance of a particular variable into two independent sources--one source which can be explained by another variable and a second source which cannot be explained by that variable. In the multivariate case, correlational analysis partitions the variance of one variable into sources that can and cannot be accounted for by combinations of other variables.

In a composite test, intercorrelations among pairs of subtests indicate the degree to which variance in one subtest score can be attributed to variance in the other subtest. To determine how much of the variation in one subtest score in a composite test is due to variation in all the other subtests, multiple correlation techniques must be used (McNemar, 1969, Chap. 11). If the other subtests each correlate zero with the subtest in question, then there is no common variance. This situation is rare in the area of psychological testing, however, because mental abilities tend to be interrelated. Generally subtests do correlate with each other to some extent, making the degree of their uniqueness difficult to determine. The procedures described in this paper apply in such cases.

As noted above, multiple correlation techniques are appropriate for determining the degree to which one subtest score can be predicted from an optimally weighted combination of the rest of the components of the test. The multiple R therefore represents the *maximum* degree to which one test component is correlated with the remaining components. The squared value of this statistic, the coefficient of multiple determination (R^2), indicates the proportion of variance in the one subtest that is dependent upon, associated with, or predicted by the other subtests. Since the scores for the other subtests are optimally weighted, R^2 indicates the maximum possible proportion of common variance.

Since subtest variance has been defined as the sum of common, unique, and error variance, the unique variance is that portion of the variance remaining after the proportions of variance attributable to common and error variance have been determined. Uniqueness analysis consists of a procedure for estimating the proportion of variance common among subtests and subtracting this from the reliable (non-error) variance of the subtest under study. A formal presentation of this procedure was given by Flanagan (1959).

UNIQUENESS ANALYSIS

Flanagan's uniqueness coefficient represents the proportion of a subtest's total variance that is both unique (i.e., free of overlap with variance measured by any other subtest in the composite) and not attributable to chance. It is defined by the formula:

$$U_i^2 = r_{ii} - \frac{R_{ic}^2}{r_{cc}} \qquad (1)$$

where U_i^2 = the uniqueness coefficient for variable i

r_{ii} = the reliability coefficient for variable i

R_{ic} = the multiple correlation (cross-validated) of variable i with the optimally weighted composite (c) of the rest of the tests

r_{cc} = the reliability of the weighted composite of the independent variables (rest of the subtests)

Equation 1 calls for a cross-validated multiple correlation coefficient (R^2_{ic}). The procedure for optimally weighting subtests in calculating the multiple R capitalizes on chance relationships in the particular sample from which the scores were obtained. In maximizing the multiple correlation, these procedures take advantage of any correlated errors or specific variations in the particular sample studied giving spuriously high estimates of the correlations among the subtests and of the multiple correlation. The degree to which chance affects the multiple correlation depends inversely on the size of the sample studied and, where applicable, on the number of variables from which a smaller variable set is to be selected. As a result of such sampling fluctuations, applying the weight determined from one sample to predict the same criterion in a new sample will result in a lower multiple correlation between predictors and criterion in the new sample. As sample size increases, sampling errors among correlations decrease, and the multiple R is less affected by capitalization on chance. This tendency for the multiple correlation to decrease as the sample size grows larger is called shrinkage.

In order to obtain values of the multiple correlation coefficient which are stable and generalizable across samples, the weights should be cross-validated. This procedure involves the application of weights or regression equations found in one sample to predictions in another sample to check on their general applicability and amount of shrinkage in the obtained multiple R.

Another method for correcting inflated multiple correlations involves the application of a shrinkage formula such as that developed by Wherry (1939). The Wherry shrinkage formula[1] results in an estimate of the multiple R for an infinitely large sample from the value obtained on a sample of specified size. When there is preselection among variables this shrinkage formula may not fully correct an inflated multiple R, but when there is no preselection among variables (all sampling error is due to selection of persons, not of variables), the formula gives an unbiased estimate of the multiple correlation and is an appropriate alternative to cross-validation. The corrected correlation coefficient represents the maximum degree to which the composite is related to the remaining subtest (variable i); the variance due to sampling errors is eliminated from the result.

Flanagan's procedure for estimating the proportion of variance common to one subtest and the combination of the other subtests corrects the estimate given by the multiple coefficient of determination (R^2) for the unreliability in the independent variables (in this case, the other subtests). The estimate is corrected for attenuation due to errors of measurement in the composite instrument. The resulting squared correlation indicates the maximum proportion of common variance possible if the composite variables had all been measured with perfect reliability. The

$$\frac{R^2_{ic}}{r_{cc}}$$

term in Equation 1 therefore represents the proportion of true score variance in the composite measure common to the observed score variance in variable i. The correction for attenuation is applied in this case because what is of interest is the degree of overlap between true performance on the composite of other subtests and actual performance on the remaining subtest.

---

1 The Wherry shrinkage formula is:

$$_sR^2 = 1 - \frac{N-1}{N-n-1}\left(1 - R^2_{ic}\right)$$

where N = number of cases
n = number of independent variables

The statistic needed to make the correction for attenuation, r_{cc}, represents the reliability of the weighted composite of the rest of the subtests. Procedures for obtaining such reliability estimates are given in Martin (1975), which also discussed different conceptual interpretations of reliability and the three most common procedures for obtaining reliability coefficients. Uniqueness analysis does not require use of a specific type of subtest reliability estimate; the test analyst should be aware of the different assumptions underlying the various reliability techniques and be careful to apply procedures appropriate for the measurement instrument being studied to obtain reliability estimates used in uniqueness analysis.

The reliability coefficient, r_{ii}, is the proportion of true score or non-error variance in variable i. The proportion of variance due to errors of measurement is $1-r_{ii}$. Subtracting the estimate of overlapping variance,

$$\frac{R^2_{ic}}{r_{cc}},$$

from the proportion of non-error variance indicates the proportion of variance in variable i that is both unique and not attributable to chance.

Equation 1 therefore partials out both the error variance in variable i, the particular subtest being studied (by using the reliability estimate r_{ii}), and the error variance in the estimate of common variance (by correcting the R^2_{ic} value for both sampling errors and errors of measurement in the composite). Flanagan (1964) notes, and it is apparent from Equation 1, that underestimation of the subtest reliabilities can result in negative values of U^2. This effect is a result not only of a smaller r_{ii} value from which to subtract the estimate of overlapping variance, but also of a spuriously high R^2_{ic} value when corrected for attenuation by an underestimate of the reliability of the composite.

Interpretation of the Uniqueness Coefficient

According to Flanagan (1964):

> ...a test is making a unique contribution to the battery if the multiple correlation (preferably cross-validated) between the test and an optimally weighted composite of all the rest of the tests is significantly less than the maximum correlation consistant with the reliability of the variables correlated. (p. 2-59).

"Significantly" is not defined in terms of statistical probability, but in practical terms. As the value of R^2_{ic}, consistent with the reliability of the weighted composite, r_{cc}, approaches the reliability of variable i, r_{ii}, the value of U^2_i decreases. As noted above, U^2_i may even be negative if underestimates of reliability are used, although theoretically zero is the lower bound (i.e., the subtest in question is making no unique contribution to test variance). Any positive value of U^2_i is therefore directly interpretable as the proportion of unique variance in that subtest (variable i).

Flanagan discussed the significance of U^2_i values in practical, operational terms. Any value of U^2_i greater than zero indicates that the subtest is making a unique contribution to the total test score. Practically, the issue of the degree to which a subtest uniquely contributes to the total test score is related to the predictive value of the total test. If the subtest is correlated with the criterion, any additional information in the test score (unique variance) related to the criterion acts to increase the validity of the total test. The degree to which each subtest potentially contributes to the total test validity is given by U^2_i. Since U^2_i represents the proportion of reliable

10

variance unique to the subtest, it also can be interpreted as the maximum degree to which the subtest's unique variance can account for variance in a pure criterion measure of its unique function. Since the square of the correlation coefficient between two variables indicates the proportion of their common variance, U_i^2 can be interpreted as a squared correlation coefficient. Therefore, U_i is an estimate of the maximum possible partial validity coefficient of a subtest with a pure criterion measure of the test's unique function when the test is used in the given test battery or a linear composite.

UNIQUENESS ANALYSIS OF TEST 500

Flanagan's procedure for uniqueness analysis was applied to the four series of Test 500 administered competitively in FY 75. Each series consists of five subtests designed to measure different abilities. Four of the abilities are measured by two types of test items, and one ability is measured by only one item type. Although alternate-forms reliability estimates indicate that the subtests in different series measure the same abilities (Martin, 1975), separate uniqueness analyses of each series were carried out in this initial effort.

Internal consistency (KR-20) reliability coefficients were used in the analysis under the assumption that errors of measurement resulted from lack of internal consistency in item sampling. These reliability coefficients were also used in estimating the reliability of the optimally weighted composite of the four subtests.

Multiple correlation coefficients obtained on large samples (e.g. greater than 1,000 cases) are relatively unbiased population estimates, particularly when the number of variables in the composite is small. It is obvious from the formula that in such situations the Wherry correction of the obtained multiple correlation coefficient is generally beyond the second decimal place. The R_{1c}^2 values obtained from the Test 500 data are unbiased to the fifth decimal place, making correction for shrinkage unnecessary. It is doubtful that any large error resulted from using the uncorrected R_{1c}^2 values since the sample sizes were so large and far exceeded the number of variables and the multiple correlations coefficients were high enough to represent nonchance relationships.

Tables 1 through 4 give the results of the uniqueness analyses by test series. The proportion of variation attributable to error, overlapping, and unique variance are given. Also presented are the potential unique validity coefficients for each test part. Since the five subtests of Test 500 were designed to measure job-related abilities, this validity estimate is meaningful because it is certain that the unique ability measured by each subtest would be part of any criterion of successful job performance.

REFERENCES

Flanagan, J. C. Flanagan Aptitude Classification Tests: Technical Report. Chicago: Science Research Associates, 1959.

Flanagan, J. C., Davis, F. B., Dailey, J. T., Shaycoft, M. F., Orr, D. B ., Goldberg, I., & Neyman, C. A., Jr. The American High-School Student. Technical Report to the U. S. Office of Education, Cooperative Research Project No. 635. Pittsburgh: Project TALENT Office, University of Pittsburgh, 1964.

Guilford, J. P., & Fruchter, B. Fundamental statistics in psychology and education (5th ed.). New York: McGraw-Hill, 1973.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

McNemar, Q. Psychological statistics (4th ed.). New York: John Wiley and Sons, Inc., 1969.

Martin, C. G. A procedure for estimating the reliability of a weighted linear composite test: the reliability of Test 500, Series 110. Washington, D. C.: U. S. Civil Service Commission, 1975.

Wherry, R. J. A new formula for predicting the shrinkage of the coefficient of multiple correlation. Annals of Mathematical Statistics, 1939, 2, 440-457.

Wing, H. The scoring system of the Professional and Administrative Career Examination. Washington, D. C.: U. S. Civil Service Commission, 1974.

TABLE 1

Uniqueness Analysis - Test 500, Series 110

Test Part	Proportion of Error Variance $1-r_{ii}$*	Multiple Correlation R_{ic}	Proportion of Overlapping Variance $\frac{R^2_{ic}}{r_{cc}}$	Proportion of Unique Variance U^2_i	Potential Unique Validity U_i
A	.146	.749	.652	.202	.449
C	.264	.739	.604	.132	.363
D	.224	.611	.425	.351	.592
E	.169	.767	.645	.186	.431
F	.265	.622	.433	.303	.550

TABLE 2

Uniqueness Analysis - Test 500, Series 120

Test Part	Proportion of Error Variance $1-r_{ii}$*	Multiple Correlation R_{ic}	Proportion of Overlapping Variance $\frac{R^2_{ic}}{r_{cc}}$	Proportion of Unique Variance U^2_i	Potential Unique Validity U_i
A	.151	.740	.645	.204	.452
C	.301	.726	.583	.116	.341
D	.195	.621	.437	.369	.607
E	.183	.757	.627	.191	.437
F	.253	.617	.425	.323	.568

*See Equation 1 for explanation of statistical terms.

TABLE 3

Uniqueness Analysis - Test 500, Series 130

Test Part	Proportion of Error Variance $1-r_{ii}$*	Multiple Correlation R_{ic}	Proportion of Overlapping Variance $\frac{R_{ic}^2}{r_{cc}}$	Proportion of Unique Variance U_i^2	Potential Unique Validity U_i
A	.167	.701	.580	.254	.504
C	.331	.675	.509	.161	.401
D	.198	.651	.475	.327	.572
E	.200	.745	.610	.191	.437
F	.230	.650	.475	.295	.543

TABLE 4

Uniqueness Analysis - Test 500, Series 140

Test Part	Proportion of Error Variance $1-r_{ii}$*	Multiple Correlation R_{ic}	Proporation of Overlapping Variance $\frac{R_{ic}^2}{r_{cc}}$	Proportion of Unique Variance U_i^2	Potential Unique Validity U_i
A	.197	.701	.585	.218	.467
C	.325	.689	.537	.139	.373
D	.199	.611	.428	.374	.612
E	.226	.758	.636	.138	.371
F	.251	.626	.448	.301	.549

*See Equation 1 for explanation of statistical terms.

GPO 908-690