

SYNTHETIC ITEM ANALYSIS:

AN APPLICATION OF SYNTHETIC VALIDITY TO ITEM ANALYSIS

by

Jerome Lehnus

A Dissertation Submitted to the Faculty of the Graduate
School of Loyola University of Chicago in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

June

1976

ACKNOWLEDGMENTS

The author wishes to thank his advisor and chairman of his committee, Dr. Jack Kavanagh, and other members of his committee, Dr. Samuel Mayo, Dr. Joy Rogers, and Dr. Donald Shepherd for their assistance in this dissertation project.

VITA

The author, Jerome David Lehnus, is the son of Walter Lehnus and Helen (Winn) Lehnus. He was born July 11, 1943, in Temple, Texas.

He attended the public high school in Lyons, Kansas and received his undergraduate education at St. Benedict's College in Atchison, Kansas and at the University of Kansas at Lawrence. He received the degree of Bachelor of Science from the School of Education of the University of Kansas in June, 1965.

He served in the Peace Corps in Colombia from the summer of 1965 to the summer of 1967 where he taught physics at the Universidad Pedagogica y Tecnologica de Colombia and acted as Science Coordinator for Peace Corps projects in Colombia.

Since his return to the United States, he has taught high school mathematics in Kansas, New Mexico, and Chicago, Illinois. He has lectured in statistics at Loyola University and in educational measurement at Northeastern University of Illinois.

He received the degree of Master of Education from Loyola University in Chicago in June, 1973.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITAiii
TABLE OF CONTENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
REVIEW OF THE LITERATURE	6
Item Analysis	7
Synthetic Validity	17
PROCEDURE	26
The Representation of Factors & Items by Vectors	27
Translating the Vector Model to Familiar Test Statistics	35
Item Selection: Internal Criterion	45
Item Selection: External Criteria	49
Construction of a Validation Criterion	50
Validation	51
RESULTS	54
DISCUSSION	61
Interpretation of Results	61
Application	70
BIBLIOGRAPHY	74
APPENDIX: The Fortran Program	76

LIST OF TABLES

Table	Page
1. Percentile ranks used to create "normal" distributions, equivalent z-scores and "typical" angles	29
2. Equivalent "100-item reliability," single-item "reliability" and "typical" angle	33
3. Sample problem: equivalent percentile ranks, z-scores, X- and Y-values.	41
4. Sample problem: sums of all possible X- and Y-values	41
5. Sample problem: sums of X- and Y-values transformed to represent item difficulty	43
6. Results: "reliability" related to validity	54
7. Results: overall direction of items related to validity.	56
8. Results: relative factor importance related to validity.	58
9. Results: correlation of factors related to validity ("reliability" = 0.90).	59
10. Results: correlation of factors related to validity ("reliability" = 0.99375).	59
11. Comparison: cosine of angle between overall direction of items and validation criterion with ratio of validity to optimum validity	63
12. Comparison: cosine of angle between overall direction of items and validation criterion resulting from changing ratio of factor importance with ratio of validity to optimum validity	65

LIST OF FIGURES

Figure	Page
1. Factor structure of test as proposed by Mosier	13
2. Example: expectancy chart used by Guion	20
3. Arrangement of item-vectors in a single plane	29
4. Arrangement of item-vector planes	30
5. Graph: "reliability" related to validity.	55
6. Graph: overall direction of items related to validity	57
7. Graph: relative factor importance related to validity	58
8. Graph: intercorrelation of factors related to validity	60
9. Venn diagram: possible relation of examination, criterion, and true job performance.	68

CHAPTER I

INTRODUCTION

The Factor Structure of Jobs

Tests, whether in education or in business, are used for a variety of purposes. One purpose is to predict the success of individuals in particular endeavors. For example, college entrance examinations are used to predict success in college. A test used in business to screen a job applicant is a measure of the applicant's probable success in that job.

Most jobs call for a variety of traits or abilities which individuals have not only in different degrees absolutely, but also in different degrees proportionally. A secretary may be required, among other things, to compose routine letters and to type them. While some individuals may be highly qualified in both of these skills and others in neither, there are those who are better qualified in one but not the other. In psychological jargon, the ability to perform a particular job consists of several factors.

The effectiveness of a selection process is limited to the degree to which it is sensitive to all of the factors that affect job performance and that exist in the

applicant population in varying amounts. Assuming that the effects of these factors are additive and that there is a linear relation between the effect of a factor and its measure, the selection process must weight each of these factors in proportion to its relative contribution to job success.

Test Validity

Ideally, the selection process consists of giving to the job applicant a test which yields a single score. That score is monotonically if not linearly related to the likelihood that the applicant will perform his job at acceptable levels. That is, applicants who receive higher scores on the test should be better workers. Tests devised up to now do not fit this criterion. Sometimes an individual with a certain score may become a better worker than another individual with a higher score. The frequency and magnitude of such reversals is indicated by the validity of the test; the more frequent and greater the reversals, the less valid the test.

In general, the validity of a selective test is defined as a correlation coefficient of the test score with some criterion of job success, such as a supervisory rating.

Test Construction

There are many procedures for constructing tests. Many follow the pattern of selecting a set of questions or items, trying them on a sample, and subjecting the items to an analysis to determine which are effectively discriminating in the desired way. Items found to be deficient are eliminated or altered.

Appropriate discrimination may be determined by comparing item statistics with the whole set of items or with some external criterion. For job applicant tests, the obvious criteria are supervisory ratings of persons hired. However, supervisory ratings are not generally regarded as adequately reliable criteria.¹ Problems with supervisory ratings as criteria for validating tests produced the invention of synthetic validity.

Synthetic Validity

Synthetic validity estimates the validity of a test with respect to job success by measuring the validity of the test with respect to each of the factors or "job ele-

¹Edwin E. Ghiselli, "The Generalization of Validity," Personnel Psychology, XII (Autumn, 1959), p. 399; Wayne K. Kirchner and Donald J. Reisberg, "Differences between Better and Less-Effective Supervisors in Appraisal of Subordinates," Personnel Psychology, XV (Autumn, 1962), p. 302; Bernard M. Bass, "Further Evidence on the Dynamic Character of Criteria," Personnel Psychology, XV (Spring, 1962), p. 93 ff.

ments" and by estimating the relative importance of these factors; the synthetic validity is a function of the test-factor validities and the relative importance of the factors.

An advantage of synthetic validity is that the process of validating a test against a population different from the sample initially used is simplified. If the same factors are involved, the relative importance of these factors must be estimated, but the test need not be tried again to determine the test-factor validity. Ernest Primoff suggests that the estimation of relative importance of these factors may be more reliable than the usual criteria, supervisory ratings.²

Synthetic Item Analysis

The technique of synthetic validity can be applied to item analysis. If a test is designed to measure potential in a certain job and more than one measurable factor contributing to that potential can be identified, then each of these factors can be treated as an external criterion against which to correlate the item. A simple process would be to assign an index of discrimination to each item based on the weighted average of its criterion correlations.

²Ernest S. Primoff, "The J-coefficient Approach to Jobs and Tests," Personnel Administration, XX (May-June, 1957), p. 36.

If this technique were as effective as item analysis based on item-whole test correlations or based on a single external criterion, it would eliminate the tendency to produce homogeneous tests and the necessity of trying items on different groups of workers and obtaining supervisory ratings for these workers. The quality of the criterion, supervisory judgment, would be improved.

Objective of This Study

The objective of this study is to show that, under certain realistic circumstances, a test constructed by using synthetic item analysis is at least as valid as one constructed by correlating item scores with whole test scores. The demonstration will use hypothetical data.

CHAPTER II

REVIEW OF THE LITERATURE

This paper proposes to apply the principle of synthetic validity to item analysis. Relevant development of item analysis and of synthetic validity will be discussed separately.

ITEM ANALYSIS

A concept central to item analysis is that of discrimination. Frequently, the purpose of testing is to discriminate between two kinds of people. For example, Binet's intelligence test was developed for the purpose of discriminating between children who would profit from schooling and those who would not.

The idea of discrimination need not be limited to two categories. Tests which report results in terms of stanines place subjects in nine categories. Theoretically, categories which are ordinal can be subdivided ad infinitum, yielding an infinite number of infinitesimal categories. This suggests that measurement, even on a continuous scale, is a process of discrimination.

Two statistics are pertinent to the discriminating power of test items: the index of difficulty and the index of discrimination.

The index of difficulty of an item is the proportion of individuals in the sample that answer the item correctly. Various measurement theorists have shown that, where guessing is not an important factor and certain other assumptions are met, a test will discriminate best if the level of difficulty for all items is 0.50.³ That is, the number of ordinal categories into which a very large sam-

ple can be sorted will be maximized if the item difficulty is 0.50.

One assumption critical to the argument favoring a 0.50 level of difficulty is that the inter-item correlations are low. If the items are such that, if a person can do one, he can do them all, the number of possible categories would be increased by spreading the level of difficulty. Sten Henrysson has presented an illustrative example:

Consider a group of 10 items to be used with 100 examinees. If all items were perfectly correlated (and thus perfectly reliable), the number of discriminations made by 10 items at 50 percent difficulty level would be identical with the number of discriminations between persons made by 1 item of 50 percent difficulty. This number of discriminations between persons is 2,500, since all the best 50 students are discriminated from the other 50 students ($50 \cdot 50 = 2,500$). But if the 10 items are spread at difficulty intervals of 9.09 percent from 9.09 percent to 90.90 percent, 4,562 discriminations could be made. The latter arrangement would be optimal for 10 items under the circumstances specified.⁴

³M. W. Richardson, "Notes on the Rationale of Item Analysis," Psychometrika, I (1936), p. 74; Lee J. Cronbach and Willard G. Warrington, "Efficiency of Multiple-Choice Tests as a Function of Spread of Item Difficulties," Psychometrika, XVII (June, 1952), p. 147; Frederic M. Lord, "The Relation of the Reliability of Multiple-Choice Tests to the Distribution of Item Difficulties," Psychometrika, XVII (June, 1952), p. 181 ff.

⁴Sten Henrysson, "Gathering, Analyzing, and Using Data on Test Items," in Educational Measurement, ed. by Robert L. Thorndike (Washington, D. C.: American Council on Education, 1971), pp. 151-52.

Richardson has shown that, if the purpose of a test is to dichotomize a population, the test will be most effective if the level of difficulty corresponds to the proportion of the population in the lower category.⁵ For example, if a selection instrument is to select the best fifteen percent of a population, the items in the instrument should be at a level of difficulty of 0.85.

In practical situations, it is generally recommended that the item difficulties be greater than 0.20 and less than 0.80 and center about 0.50.⁶ For true-false and multiple-choice tests, these figures are adjusted upward to compensate for the "guessing" effect.

The index of discrimination is some measure of association which compares the pattern of discrimination of an item with some criterion: either the whole test score or some external criterion. The most obvious measure of association is a correlation of item response to the criterion. If item responses are scored either right or wrong, the correlation will logically be a biserial or a point-biserial. If the criterion is dichotomous, a tetrachoric or phi-coefficient is indicated.

⁵M. W. Richardson, "The Relation between the Difficulty and the Differential Validity of a Test," Psychometrika, I (June, 1936), p. 47 ff.

⁶Jum C. Nunnally, Educational Measurement and Evaluation (New York: McGraw-Hill Book Company, 1972), p. 188; Henrysson, "Gathering, Analyzing, and Using Data," p. 144.

Frequently, it is recommended that the criterion be used to divide the sample into three parts: a lower, a middle, and an upper group. Given certain assumptions, Kelley has shown that item discrimination can be most efficiently estimated if the upper and lower groups each contain twenty-seven percent of the sample scores.⁷ A simple discrimination statistic using upper and lower groups consists simply of the differences of the number of correct responses made to an item by members of the upper group less correct responses to the item by members of the lower group.⁸

Some of the procedures mentioned above are favored over others on the grounds that they tend to select items which have a level of difficulty near 0.50. The index of discrimination which results from subtracting the number of correct responses of a lower group from those of an upper group, for example, is clearly biased against very easy and very difficult items.

⁷Robert L. Ebel, Essentials of Educational Measurement (Englewood Cliffs: Prentice-Hall, Inc., 1972) p. 386, citing Truman L. Kelley, "The Selection of Upper and Lower Groups for the Validation of Test Items," Journal of Educational Psychology, XXX (1939), pp. 17-24.

⁸Robert L. Ebel, Essentials of Educational Measurement (Englewood Cliffs: Prentice-Hall, Inc., 1972) p. 388, citing A. Pemberton Johnson, "Notes on Suggested Index of Item Validity: The U-L Index," Journal of Educational Psychology, LXII (1951), pp. 499-504.

An empirical comparison of a variety of indices of discrimination suggests that they yield essentially the same information.⁹

Procedures which use the whole test scores as a criterion are justified by the assumption that the test constructor has selected valid items on the whole, even though some items may be defective. Defective items are identified through their inconsistency with the test constructor's overall good judgment. Further justification of this procedure is based on the interrelationship of reliability and validity.

A test is said to be reliable if it measures something consistently. The measure of reliability is generally a correlation coefficient. The correlation may be between sets of scores obtained by giving the same test to a group of individuals on different occasions, by giving alternate forms of a test to the group, or by splitting a test into two equivalent halves and comparing the scores on the two halves. In the latter case, the resulting correlation is corrected for the decreased number of items. Richardson has shown that the reliability of a test is a function of the intercorrelations of the items in the test and that item analysis increases the reliability of the

⁹Max D. Englehart, "A Comparison of Several Item Discrimination Indices," Journal of Educational Measurement, II (June, 1965), p. 69 ff.

test by eliminating the items which have lowest intercorrelations with the other items.¹⁰ In this sense, the test is made more homogeneous.

A test is said to be valid if it measures what it purports to measure and does not measure things incongruent with what it purports to measure. Logically, if a test is not a consistent measure of itself, it cannot be a consistent measure of anything else. The square root of the reliability of a test is an upper limit of its validity; if a test is not reliable, it cannot be valid.

Procedures which use the whole test score as a criterion, then, are also justified by the fact that they do increase reliability. While this does not necessarily raise the validity of the test, it at least raises the upper limit of the validity. The test is given the opportunity to be more valid.

Charles Mosier has presented a model for tests and factors which illustrates one of the problems for item-whole test item analysis. If a test measures more than one psychological factor, which it almost certainly must, these factors may be thought of as vectors. To simplify the argument, suppose that only two factors are involved, as illustrated in figure 1:

¹⁰Richardson, "Notes on the Rationale of Item Analysis," p. 74.

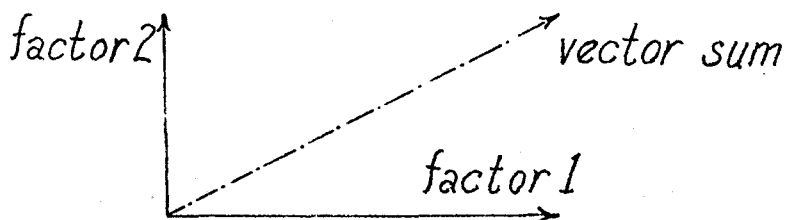


figure 1

Suppose that these factors are chosen in such a way that factor 1 is congruent with the purpose of the test and that factor 2 is orthogonal to factor 1 and inconsistent with the purpose of the test. With respect to the purpose of the test, factor 2 represents systematic error. Item analysis will select items whose factor structure resembles that of the whole test. That is, it will select items whose vectors are aligned with the vector sum. Items with some systematic error are preferred to those parallel to the true purpose of the test. If factor 2 were large relative to factor 1, item analysis might actually make the test less valid, though more reliable.¹¹

Henrysson comments that if a test is intended to measure a variety of factors, item analysis may make the test less valid by making it too narrow to have content validity.¹²

¹¹Charles I. Mosier, "A Note on Item Analysis and the Criterion of Internal Consistency," Psychometrika, I (December, 1936), p. 275 ff.

¹²Henrysson, "Gathering, Analyzing, and Using Data on Test Items," p. 154.

Most of the procedures which have emerged over the years have been developed primarily with computational convenience rather than statistical theory in mind.¹³ Robert Ebel has pointed out that the advantages of using internal criteria are those of convenience: relevant external criteria may be difficult to find and whole test scores are always available.¹⁴ The following study illustrates this point.

David Ryans developed two tests from a common set of items using internal criteria for one and external criteria for the other. The test ostensibly measured teachers' professional knowledge. The external criterion was supervisory (principal's) ratings for job performance. Ryans noted that the external criterion probably included various factors other than teachers' professional knowledge. That is, the external criterion was not altogether pertinent. He found that the test resulting from the use of internal criteria was more homogeneous than that resulting from the use of an external criterion.¹⁵ This

¹³Ibid., p. 145.

¹⁴Robert L. Ebel, Measuring Educational Achievement (Englewood Cliffs: Prentice-Hall, Inc., 1965), p. 357.

¹⁵David G. Ryans, "The Results of Internal Consistency and External Validation Procedures Applied in the Analysis of Test Items Measuring Professional Information," Educational and Psychological Measurement (1951), p. 558.

means that the use of external criteria in item analysis cannot be expected to produce as homogeneous (reliable) a test as the use of internal criteria.

In some circumstances, the use of external criteria does not significantly improve validity, either. David Hasson selected items from the Otis-Lennon Mental Ability Test on the basis of the total test score and on the basis of a criterion measure, the Metropolitan Achievement Test. He found no significant difference in the predictive ability of the resulting tests.¹⁶

Henrysson has suggested that the increased availability of computers will allow more statistically sophisticated and theoretically justifiable procedures to take precedence over computational convenience.¹⁷ The following study seems to support this prediction.

John Fossum developed two tests from a common set of items using external criteria in both cases. In one case, he used a regression procedure, selecting items "so that at each iteration the item selected is the one leading to the largest increase in correlation."¹⁸ He called

¹⁶David J. Hasson, "An Evaluation of Two Methods of Test Item Selection," Dissertation Abstracts, Vol. 32A 6200-A.

¹⁷Henrysson, "Gathering, Analyzing, and Using Data on Test Items," pp. 155-56.

¹⁸John A. Fossum, "An Application of Techniques to

this procedure the "sequential nominator method." The other test was constructed by selecting items in descending order of their correlations with the criteria. An equal number of items were selected by both methods so that the size of the resulting tests would not influence their relative validities. The former method produced the more valid test. He concluded that "If the item inter-correlation matrix is stable across samples, then the sequential method is superior to one which does not consider intercorrelations."¹⁹ This conclusion is qualified: "If the intercorrelations are low, there is little advantage in using the more complex sequential nominator method."²⁰

Shorten Tests and Increase Validity," Journal of Applied Psychology, LVII (February, 1973), p. 90.

¹⁹Ibid., p. 92.

²⁰Ibid.

SYNTHETIC VALIDITY

The term "synthetic validity" was introduced by
C. H. Lawshe

to denote the inferring of validity in a specific situation. The concept is similar to that involved when the time study engineer establishes standard times for new operations, purely on an a priori basis through the use of "synthetic times" for the various elements constituting the operation.²¹

The concept is more specifically related to jobs by Michael Balma, who defines synthetic validity as

the inferring of validity in a specific situation from a logical analysis of the jobs into their elements, a determination of test validity for these elements, and a combination of elemental validities into a whole.²²

Edwin Ghiselli presents as the genesis of synthetic validity the fact that validities for the same test/job in different locations show little or no agreement. He reports that the variance of validity coefficients is greater than could be accounted for by random variation alone. Two reasons are offered for this phenomenon: (1) the criteria used to establish the validity correlations are not stable and (2) the "fact that the same job

²¹C. H. Lawshe and Martin D. Steinberg, "An Exploratory Investigation of Clerical Jobs," Personnel Psychology, VIII (1955), p. 291.

²²Michael J. Balma, "The Concept of Synthetic Validity," Personnel Psychology, XII (Autumn, 1959), p. 399.

in two different establishments is not in fact the same job," i.e., jobs of the same title vary in their requisite duties and abilities from one location to another.²³

Ernest Primoff points out that the use of synthetic validity allows the estimation of validity, and therefore the selection of tests, for jobs in which there are too few individuals to permit validation in the usual way and for new jobs for which no incumbent workers are available for traditional validation studies.²⁴

The process of synthetic validity may be divided into three parts: (1) the identification of the knowledges, skills, and personality traits which contribute to the performance of a job and the determination of their relative importance, (2) the determination of the relationship of test scores to the skills and so forth that are identified, and (3) the combination of these two types of information into a single estimator of an individual's job potential.

To show the feasibility of the first two parts of this procedure, Lawshe and Steinberg investigated the relationship of parts of clerical workers' jobs and the workers' scores on related parts of the Purdue Clerical Adaptability Test. They found that workers who were frequently

²³Edwin E. Ghiselli, "The Generalization of Validity," Personnel Psychology, XII (Autumn, 1959), p. 399.

²⁴Ernest S. Primoff, "The J-Coefficient Approach to Jobs and Tests," Personnel Administration, XX (May-June, 1957), p. 39.

called upon to perform a test-related task scored higher on relevant parts of the test. For example, workers frequently called upon to perform arithmetic computations scored above the median on those parts of the test calling for arithmetic computation.²⁵

Robert Guion, in order to demonstrate the feasibility of synthetic validity, used synthetic validity procedures and regression procedures to select tests for personnel hiring.²⁶ The data indicated that the synthetic validity procedures selected tests which more accurately predicted job success. The procedure of his study is as follows:

Job elements were culled from detailed descriptions of various jobs in a small company. Extensive lists of elemental tasks and abilities were prepared and grouped into seven categories or factors, such as "salesmanship," "creative business judgment," "routine judgment," and so forth. The development of these seven categories was based on the subjective judgment of Guion and of the company executives.

Two executives ranked employees with respect to each of the factors. Only employees with whom the execu-

²⁵Lawshe and Steinbert, "An Exploratory Investigation," pp. 291-97.

²⁶Robert M. Guion, "Synthetic Validity in a Small Company: A Demonstration," Personnel Psychology, XVIII

tives were familiar and whose job called for the factor in question were ranked on any particular factor. The ranks were converted to normalized scale values for the purpose of determining interrater reliabilities and interfactor correlations.

A battery of tests, producing nineteen scores was given to all of the employees. These scores were correlated to the rankings on the seven factors. It was arbitrarily decided that the two tests which best correlated with each factor be used as the predictor of that factor. Expectancy charts, such as the one shown below, were developed for each category and its related subtests.

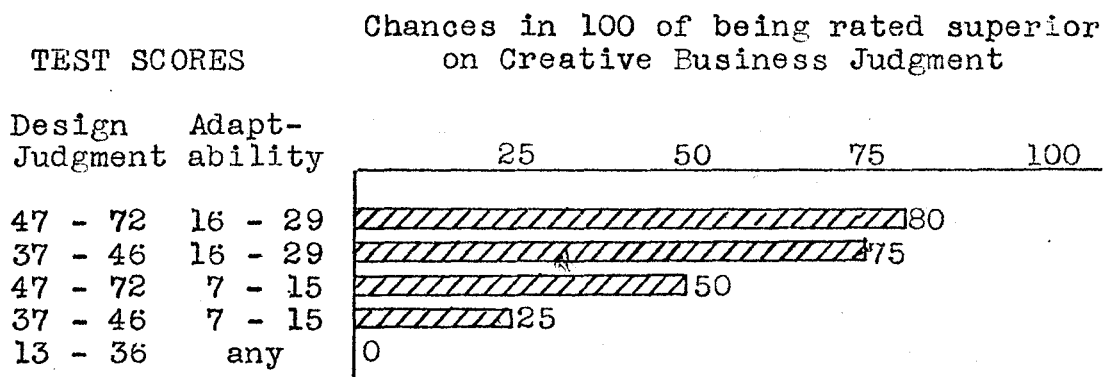


figure 2

The synthetic "validities" were applied to hiring by giving applicants tests relevant to the factors required by the position for which they were applying. For each category, the probability that the applicant would be

judged superior was determined and this probability was converted to an integer index. An applicant's "score" was the sum of the indices of the factors relevant to his prospective position. Applicants' scores were used to rank them in order of their most probable superiority in their position.

Guion compared the success of this procedure in hiring thirteen new employees to that which would have resulted from the selection of tests by multiple regression using a single job performance rating as a criterion. He found that the synthetic validity technique picked "superior" workers 76% of the time, while the multiple regression technique picked "superior" workers only 46% of the time. Because of the small number involved, this difference is not statistically significant.

Ernest Primoff has proposed a different approach to synthetic validity, which he calls the J-coefficient. It differs from Guion's treatment in two aspects: (1) the estimation of the relative importance of job factors and (2) the estimation of test-job validity.

For the estimation of relative job factor importance, Primoff's method relies on the subjective judgment of a panel of experts who are familiar with the job being analyzed. These experts are likely to be persons who have experience working at the job itself or who have experience supervising the job. Each expert is asked to rate each

job element or factor on a three point scale. An item is rated 0 if it is not important, 1 if it is moderately important, and 2 if it is of the utmost importance. For each item, the ratings of all of the raters in the panel are added. Thus, if ten raters are used, the rating for a particular element could have any integral value from 0 to 20. These totals are used to determine the relative importance of each element rated; the absolute value of the totals do not enter into subsequent calculations. Because only relative values are used, the size of the rating group and any tendencies of the group to rate toward one end of the scale do not affect subsequent calculations.²⁷

Primoff argues that the use of several raters yields precise and reliable ratings. Furthermore, this approach to analyzing job requirements has advantages over approaches which incorporate the rating of actual workers: (1) if a rater rates an ability with respect to a job rather than a worker, he is not so likely to be affected by personal bias; (2) the rating of job elements is not dependent upon variance in the ability among workers present; (3) since workers can be used as raters, it is easier to find a large number of raters who are intimately familiar with the job.²⁸

²⁷Primoff, "The J-Coefficient Approach," p. 36.

The J-coefficient is an estimate of the criterion validity of a test with respect to a job. The usual procedure to establish the criterion validity of a test which is intended to select workers is to compute the product-moment correlation of the test scores with supervisory ratings of job performance. The mathematical formula for this correlation is

$$r = \frac{\sum x_i y_i}{N \sqrt{(\sum x_i^2 / N)(\sum y_i^2 / N)}} \quad (1)$$

where x_i is the i th person's deviation test score, y_i is his deviation criterion score, and N is the number of persons in the validation sample. The criterion score might be a supervisory rating, such as the normalized ratings described in Guion's study, mentioned above. Generally, the statistical treatment of this type of correlation assumes that both variables are normally distributed and homoscedastic and that one variable is a linear function of the other.

If a test measures more than one job factor and if z_{ik} denotes a standardized supervisory rating of the i th worker on the k th job element, regression equations may be written which estimate the test score in terms of job element ratings:

$$\hat{x}_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_n z_{in} \quad (2)$$

Similarly, a regression equation could be written which would predict y_1 ; call the estimate \hat{y}_1 . The estimated validity coefficient could be computed as

$$J = \frac{\sum \hat{x}_i \hat{y}_i}{N \sqrt{(\sum \hat{x}_i / N)(\sum \hat{y}_i / N)}} \quad (3)$$

Using matrix algebra, Primoff shows that this equation is equivalent to

$$J = \sum_k \beta_k r_{yk} \quad (4)$$

where β_k is the regression coefficient in equation (2) and r_{yk} is the product-moment correlation of the k^{th} job element rating with the overall supervisory rating. The disappearance of the denominator assumes that the list of job elements is virtually complete and that multiple correlations of the job elements to the test and to the supervisory ratings are near unity.²⁹

In practice, r_{yk} is derived from intercorrelations of the job element ratings and relative importance "weights" of the job elements assigned by job experts. If w_j denotes the weight assigned by the job experts to the j^{th} element and r_{jk} is the correlation of the j^{th} element and the k^{th} element as determined by the ratings, the derived corre-

²⁹Ernest S. Primoff, Basic Formulae for the J-coefficient to Select Tests by Job Analysis Requirements (Washington, D. C.: Test Development Section, United States Civil Service Commission, 1955)

lation is given by³⁰

$$r_{yk} = \frac{\sum_j r_{jk} w_j}{\sqrt{\sum_j w_j (\sum_j r_{jk} w_j)}} \quad (5)$$

Primoff indicates that the J-coefficient has been used successfully in the development of selection batteries.³¹ Dane Selby found it feasible in "public jurisdictions which have large applicant populations." He did not find it "quick and inexpensive when compared to traditional validation studies."³²

³⁰Ibid.

³¹Primoff, "The J-Coefficient Approach," p. 34.

³²Dane Selby, The Validation of Tests Using J-Coefficient: A Feasibility Study, (Illinois: Research and Test Development, Illinois Department of Personnel, 1975), p. 3.

CHAPTER III

PROCEDURE

The procedure of this study consists of:

- 1) developing a hypothetical situation involving job factors and test items described in terms of vectors,
- 2) translating these vectors to the kind of numbers typically used as test item statistics,
- 3) selecting a set of those items according to an internal criterion,
- 4) selecting another set of items according to a technique which applies the principles of synthetic validity to item statistics,
- 5) constructing a criterion for validation from the job factor-vectors, and
- 6) validating the sets of items resulting from the different selection techniques against the validation criterion and comparing the results.

Each of these steps will be discussed in more detail in the following sections.

THE REPRESENTATION OF FACTORS & ITEMS BY VECTORS*

Job elements or factors have two salient mathematical features, their relative importance and their intercorrelation. Both of these can be represented by vectors. The relative importance of a factor is analogous to the length of the vector. The intercorrelations of factors is represented by the angle between the vectors. The product-moment correlation is equal to the cosine of the angle between the vectors.

The items also can be represented by vectors; their direction will indicate their correlation with the factors. The length of the item-vectors could be used to represent their relative weights. In this study, all of the items will be assumed to be equally weighted; the lengths of the item-vectors will be equal and therefore of no consequence. As with interfactor correlation, the correlation of an item and a factor is the cosine of the angle between them.

The procedure may best be explained by presenting a simple example. Suppose that there is a job which involves two orthogonal factors, one of which is twice as

*In this paper, vectors are not intended as mathematical proof of the hypotheses presented. They are used to facilitate understanding of the procedures that involve conventional item statistics and to aid in the construction of hypothetical statistics.

influential as the other. These are represented by the solid lines in figure 3 (p. 29). Suppose also that there is a set of items which measures these factors exclusively. That is, all of the variance in response to the items can be accounted for by the variance of the factors. Geometrically, this simply implies that the item-vectors are in the same plane as the factor-vectors.

This example will also suppose that the direction of item-vectors is normally distributed with the direction of factor 1 as the mean direction of the item-vectors. Let 80° be taken as a "typical" angle between an item-vector and the vector representing factor 1. That is, the standard deviation of the angles of the item-vectors with factor 1 will be arbitrarily set at 80° .

An approximation of a normal distribution may be obtained by finding z-scores equivalent to various percentile ranks at equivalent intervals. In this demonstration, an array of fifteen z-scores is used. These are equivalent to percentile ranks running from 3.33 to 96.67 by intervals of $1/15$. These values, multiplied by a "typical" angle, 80° , will yield "normally" distributed item-vectors.* This procedure is illustrated in the following table.

*Technically, this distribution cannot be normal; its distribution function is a step function, not a continuous curve.

PERCENTILE RANK	z-SCORE	ANGLE IN DEGREES
3	-1.83	-146.4
10	-1.28	-102.4
17	-0.97	- 77.6
23	-0.73	- 58.4
30	-0.52	- 41.6
37	-0.34	- 27.2
43	-0.17	- 13.6
50	0.00	0.0
57	0.17	13.6
63	0.34	27.2
70	0.52	41.6
77	0.73	58.4
83	0.97	77.6
90	1.28	102.4
97	1.83	146.4

table 1

The position of these item-vectors relative to the factor-vectors is illustrated below.

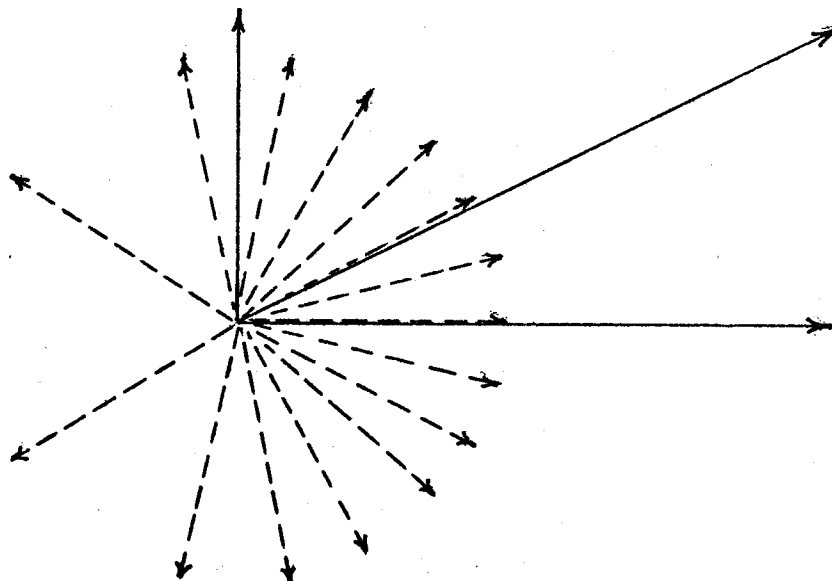


figure 3

Due to the small number of points used, this approximation to a normal distribution with mean 0 and a standard deviation of 80° is not perfect. The standard deviation of

the angles is 75.08° . Considering the arbitrariness of the selection of a "typical" angle, this discrepancy is not important.

The item-vectors are distributed not only in the plane of the two job factors but also along an error dimension. This can be imagined as having fifteen pages, each with an item-vector distribution such as that shown in figure 3, fanned out according to the angles given in table 1. That is, if the pages were bound along the line of factor 2 and their angle with factor 1 were given by table 1, the distribution of the item-vectors on those pages would be the distribution of item-vectors in the present example. Figure 4 attempts to illustrate this.

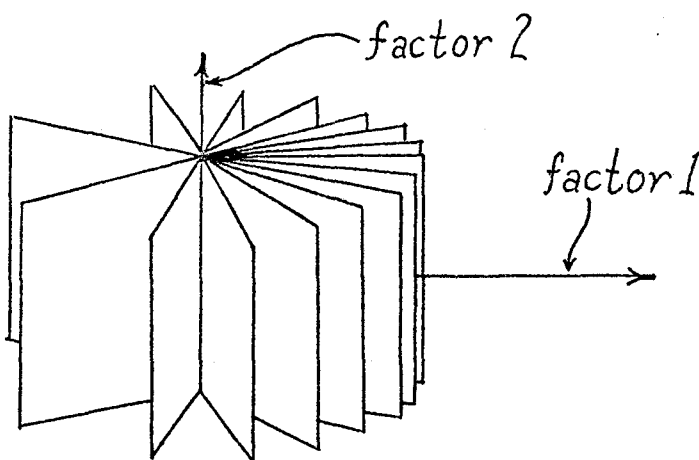


figure 4

This picture fails in that it doesn't provide for the item-vectors whose angle with factor 1 (or the horizontal) on each page exceeds 90° . Such is life.

The value of each item-vector can be represented by an ordered pair of "coordinates;" the first specifies the angle in the factor-vector plane; the second specifies the angle to the factor-vector plane.

The entire set of hypothetical items contains 225 items. From these, about 100 items will be selected. These figures are not untypical of test construction procedures. One hundred items would represent a reasonably large test, but not an uncommonly large test. Developing twice as many items as are to be eventually selected is not unusual.

There are several parameters which control the arrangement of item-vectors and factor-vectors. This paper treats four of these:

- 1) the spread of item-vectors,
- 2) the overall direction of the item-vectors,
- 3) the relative size of the factor-vectors, and
- 4) the angle between the factor-vectors.

The spread of the item-vectors is controlled by controlling the "typical" angle multiplied by various z-scores as illustrated by table 1 (p. 29). As the spread of items can be identified with reliability, the selection of a "typical" angle is identified with the selection of a realistic reliability. In this experiment, several values are posited as whole-test reliabilities. Consequent item reliabilities and angles are derived as follows:

According to the Spearman-Brown prophecy formula,

$$R = \frac{nr}{1 + (n-1)r} \quad (6)$$

where R is the reliability of the whole test, n is the number of items, and r is the reliability of each item.³³ While this formula assumes that all items are equally reliable and the items in this experiment are clearly not equally reliable, it still serves the purpose of selecting a reasonable value for a "typical" item; lack of rigor on this point does not affect the conclusions of the study.

For a test of one hundred items, equation (6) becomes

$$R = \frac{100r}{1 + 99r} \quad (7)$$

It follows that

$$r = \frac{R}{100 - 99R} \quad (8)$$

Table 2 gives the values of R used in this experiment as well as the consequent values of r and of the "typical" angle used to define the distribution of the item-vectors. The "typical" angle given is simply the inverse cosine of r .

³³Julian C. Stanley, "Reliability," in Educational Measurement, ed. by Robert L. Thorndike (Washington, D. C.: American Council on Education, 1971), p. 395.

R	r	angle
.99375	.614	52.1°
.9875	.441	63.8
.975	.281	73.7
.95	.160	80.8
.90	.083	85.2
.80	.038	87.8
.60	.015	89.1
.40	.007	89.6

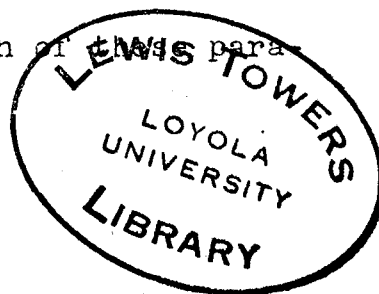
table 2

The overall direction of the item-vectors is controlled simply by adding some constant to the first coordinate (mentioned on p. 31) of each item-vector. The values used in this experiment are: -45° , -30° , -15° , 0° , 15° , 30° , and 45° . These angles are measured from factor 1 and rotation toward factor 2 is considered positive.

The relative size of the two job factor-vectors is controlled by assigning factor 2 a unit length and varying the size of factor 1. The values of factor 1's length used in this demonstration are: 0.25, 0.5, 1, 2, and 4.

The angle between the two factor-vectors is assigned the values of 90° , 80° , 70° , 60° , 50° , and 40° .

Generally, three of the parameters mentioned are held constant while the fourth assumes all of the values indicated above. The values used for each of these parameters as they are held constant are:



test "reliability".	0.90
overall direction of item-vectors	0.00°
relative size of factor-vectors (factor 1/factor 2).	2.00
angle between factor vectors.	90.0°

The various values of the angle between the factor-vectors are examined with a test "reliability" of 0.99 as well as 0.90.

TRANSLATING THE VECTOR MODEL TO FAMILIAR TEST STATISTICS

In order for any personality trait to be considered a factor in job success, it must exist in varying degrees in the worker (or applicant) population. Thus, it is reasonable to represent the job factors by random variables. In order to mimic a normal distribution, the factors will be assigned z-scores or multiples of z-scores which correspond to equal-interval percentile ranks. This is the same procedure used in distributing angles of item-vectors described on p. 28, above. The values of the variables associated with the job factors represent degrees of the trait involved, e.g., degrees of intelligence, degrees of conscientiousness, etc.

It is also reasonable to expect that persons will differ in their expected score on any item or combination of items. If a correct response is given a score of one, and an incorrect response is given a score of zero, the expected score is simply the probability that a person makes a correct response to an item.

Since the items are correlated with the job factors, as indicated by the vector model, the variable assigned to the item must be correlated to the variables assigned to the job factors. Moreover, the angles between the item-vectors and the job factor-vectors indicate specific

values for these correlations. The immediate problem, then, is to generate a random variable, S, which has specific correlations with other random variables, X and Y. The variables X and Y represent the factors.

Let X be a random variable representing factor 1, with a mean of zero and a standard deviation of σ_x . Let Y be a random variable representing factor 2, with a mean of zero and a standard deviation of σ_y . Let Z be a random variable representing error, the dimension perpendicular to the job factor plane. Let Z have a mean of zero and a standard deviation of σ_z . Let ρ_{xy} , ρ_{xs} , and ρ_{ys} denote the correlations of X with Y, X with S, and Y with S, respectively. As Z is always taken to be orthogonal to the job factors, the correlations of X with Z, ρ_{xz} , and of Y with Z, ρ_{yz} , are zero.

Let $S = X + Y + Z$. Then, by definition,

$$\sigma_s^2 = \frac{1}{n} \sum (S - \bar{S})^2 \quad (9)$$

$$= \frac{1}{n} \sum (X + Y + Z - \bar{X} - \bar{Y} - \bar{Z})^2 \quad (10)$$

$$= \frac{1}{n} \sum (X + Y + Z)^2 \quad (11)$$

$$= \frac{1}{n} \sum (X^2 + Y^2 + Z^2 + 2XY + 2XZ + 2YZ) \quad (12)$$

where n represents the number of elements of S. Since

$$\rho_{xz} = \rho_{yz} = 0,$$

$$\sigma_s^2 = \frac{1}{n} \sum (X^2 + Y^2 + Z^2 + 2XY) \quad (13)$$

$$= \sigma_x^2 + \sigma_y^2 + \sigma_z^2 + 2 \operatorname{cov}(X, Y) \quad (14)$$

$$= \sigma_x^2 + \sigma_y^2 + \sigma_z^2 + 2 \sigma_x \sigma_y \rho_{xy} \quad (15)$$

By definition,

$$\rho_{xs} = \frac{1}{n} \sum XS / \sigma_x \sigma_s \quad (16)$$

$$= \frac{1}{n} \sum X(X+Y+Z) / \sigma_x \sigma_s \quad (17)$$

$$= \frac{1}{n} \sum (X^2 + XY + XZ) / \sigma_x \sigma_s \quad (18)$$

$$= (\sigma_x^2 + \operatorname{cov}(X, Y) + 0) / \sigma_x \sigma_s \quad (19)$$

$$= (\sigma_x^2 + \sigma_x \sigma_y \rho_{xy}) / \sigma_x \sigma_s \quad (20)$$

$$= (\sigma_x + \sigma_y \rho_{xy}) / \sigma_s \quad (21)$$

Similarly,

$$\rho_{ys} = (\sigma_y + \sigma_x \rho_{xy}) / \sigma_s \quad (22)$$

Therefore,

$$\rho_{xs} / \rho_{ys} = \frac{(\sigma_x + \sigma_y \rho_{xy}) / \sigma_s}{(\sigma_y + \sigma_x \rho_{xy}) / \sigma_s} = \frac{\sigma_x + \sigma_y \rho_{xy}}{\sigma_y + \sigma_x \rho_{xy}} \quad (23)$$

$$\rho_{xs} (\sigma_y + \sigma_x \rho_{xy}) = \rho_{ys} (\sigma_x + \sigma_y \rho_{xy}) \quad (24)$$

$$\sigma_y \rho_{xs} + \sigma_x \rho_{xy} \rho_{xs} = \sigma_x \rho_{ys} + \sigma_y \rho_{xy} \rho_{ys} \quad (25)$$

$$\sigma_y (\rho_{xs} - \rho_{xy} \rho_{ys}) = \sigma_x (\rho_{ys} - \rho_{xy} \rho_{xs}) \quad (26)$$

$$\sigma_y = \sigma_x \frac{(\rho_{ys} - \rho_{xy} \rho_{xs})}{(\rho_{xs} - \rho_{xy} \rho_{ys})} \quad (27)$$

Let

$$A = \frac{\rho_{Ys} - \rho_{XY}\rho_{Xs}}{\rho_{Xs} - \rho_{XY}\rho_{Ys}} \quad (28)$$

and

$$\sigma_Y = \sigma_X A \quad (29)$$

Substituting equations (15) and (29),

$$\rho_{Xs} = \frac{\sigma_X + \sigma_X A \rho_{XY}}{\sqrt{\sigma_X^2 + \sigma_X^2 A^2 + \sigma_Z^2 + 2\sigma_X^2 A \rho_{XY}}} \quad (30)$$

Squaring both sides,

$$\rho_{Xs}^2 = \frac{\sigma_X^2 + 2\sigma_X^2 A \rho_{XY} + \sigma_X^2 A^2 \rho_{XY}^2}{\sigma_X^2 + \sigma_X^2 A^2 + \sigma_Z^2 + 2\sigma_X^2 A \rho_{XY}} \quad (31)$$

$$\rho_{Xs}^2 (\sigma_X^2 + \sigma_X^2 A^2 + \sigma_Z^2 + 2\sigma_X^2 A \rho_{XY}) = \sigma_X^2 + 2\sigma_X^2 A \rho_{XY} + \sigma_X^2 A^2 \rho_{XY}^2 \quad (32)$$

$$\sigma_X^2 \rho_{Xs}^2 (1 + 2A \rho_{XY} + A^2) + \rho_{Xs}^2 \sigma_Z^2 = \sigma_X^2 (1 + 2A \rho_{XY} + A^2 \rho_{XY}^2) \quad (33)$$

$$\rho_{Xs}^2 \sigma_Z^2 = \sigma_X^2 [(1 + 2A \rho_{XY} + A^2 \rho_{XY}^2) - \rho_{Xs}^2 (1 + 2A \rho_{XY} + A^2)] \quad (34)$$

$$\sigma_X^2 = \frac{\rho_{Xs}^2 \sigma_Z^2}{(1 + 2A \rho_{XY} + A^2 \rho_{XY}^2) - \rho_{Xs}^2 (1 + 2A \rho_{XY} + A^2)} \quad (35)$$

And from (29),

$$\sigma_Y^2 = \sigma_X^2 A^2 \quad (36)$$

Z is assigned the fifteen "normally" distributed values given in table 1. The variance of these values is 0.9168. X is assigned these values multiplied by σ_X/σ_Z where $\sigma_X = \sqrt{\sigma_X^2}$ and σ_X^2 is given by (35), provided ρ_{Xs} is positive. If ρ_{Xs} is negative, the sign of these values is changed. Y is assigned the values of Z multiplied by σ_Y/σ_Z where $\sigma_Y = \sqrt{\sigma_Y^2}$ and σ_Y^2 is given by (36),

provided ρ_{Ys} is positive. If ρ_{Ys} is negative, the sign of these values is changed.

The procedure described, obtaining the desired intercorrelations by setting appropriate relative standard deviations, fails when σ_x^2 or σ_z^2 is zero. The former case arises when the item-vector is in the same plane as the job factor-vectors. If the item-vector is in that plane, the denominator of (35) will be zero. That is, $\sigma_x^2 = 0$ if and only if

$$(1 + 2A\rho_{XY} + A^2\rho_{XY}^2) = \rho_{Xs} (1 + 2A\rho_{XY} + A^2) \quad (37)$$

The variance of X, σ_x^2 , is zero if and only if $\rho_{Xs} = 0$.

In the instance that both σ_x^2 and σ_z^2 are zero, σ_y^2 is assigned a value of one. If σ_z^2 is zero, and σ_x^2 is not zero, σ_x is assigned a value of one (or negative one) and σ_y^2 is determined by equation (36). If σ_x^2 were zero and σ_z^2 were not, equations (15) and (26) could be used to solve for σ_y^2 in terms of σ_z^2 . The latter circumstance does not arise in this particular demonstration.

In all cases described, values are assigned to the variables X, Y, and Z by multiplying the z-scores given in table 1 by appropriate scaling factors.

The variable S has 15^3 values, these being the sums of all possible combinations of the values of X, Y, and Z. It represents placement of an individual with respect to an item. In order to give it the appearance of a probability, as indicated on p. 35, it must have values be-

tween 0 and 1. It is arbitrarily decided to make S have a mean value of 0.5 and a standard deviation of 0.1. The value of 0.5 is suggested by test theorists as ideal, as noted and qualified on p. 8. The value of 0.1 as a standard deviation makes it highly unlikely that any person on any item will have an expected score less than 0 or greater than 1. These parameters are imposed upon S by a linear transformation:

$$S' = \frac{0.1}{\sigma_s} S + 0.5 \quad (38)$$

where S is the value which is the sum of X, Y, and Z and S' is the "corrected" value of S. No correction for the initial mean of S is indicated in this formula, as this is 0. The linear transformation does not affect the correlations of S (or S') with X or Y.

The entire process described so far might be made clearer by an example. Suppose there is an item whose elevation above factor 1 is 30° and whose angle with the factor plane is 40° . Suppose also that the angle between the job factor-vectors is 90° . Thus,

$$\rho_{XY} = \cos 90^\circ = 0.0 \quad (39)$$

$$\rho_{XS} = \cos 30^\circ \cdot \cos 40^\circ = 0.66 \quad (40)$$

$$\rho_{YS} = \cos(90-30)^\circ \cdot \cos 40^\circ = 0.38 \quad (41)$$

The purpose of the illustration can be just as well served using five values for the factors rather than fif-

teen. These are chosen as the z-scores corresponding to the 10th, 30th, 50th, 70th, and 90th percentiles. These values are given in table 3. The variance of these z-scores is 0.764.

PERCENTILE RANK	z-SCORE	X	Y
10	-1.28	-1.30	-0.75
30	-0.52	-0.53	-0.30
50	0.00	0.00	0.00
70	0.52	0.53	0.30
90	1.28	1.30	0.75

table 3

According to equation (26), $\sigma_x^2 = 0.79$ and $\sigma_x = 0.89$. According to equation (27), $\sigma_y^2 = 0.26$ and $\sigma_y = 0.51$. Multiplying the standard deviation of X and Y by the z-scores and dividing by the standard deviation of the z-scores yields the values given for X and Y in table 3. These are also the marginal values of table 4.

		Y				
		-0.75	-0.30	0.00	0.30	0.75
X	-1.30	-2.05	-1.60	-1.30	-1.00	-0.55
	-0.53	-1.28	-0.83	-0.53	-0.23	0.22
	0.00	-0.75	-0.30	0.00	0.30	0.75
	0.53	-0.22	0.23	0.53	0.83	1.28
	1.30	0.55	1.00	1.30	1.60	2.05

table 4

The figures in the cells of table 4 are simply the sums of the marginal values. These figures do not represent

all of the values of S. The values of S could be obtained by adding the values of Z (z-scores) to the figures in the cells. If this were done, the first cell, for example, would contain the values

$$-3.33 = -2.05 + -1.28,$$

$$-2.57 = -2.05 + -0.52,$$

$$-2.05 = -2.05 + 0.00,$$

$$-1.53 = -2.05 + 0.52,$$

$$-0.77 = -2.05 + 1.28.$$

However, the figures in table 4 do suffice to establish correlations and define expected scores for individuals of characteristics defined by X and Y.

The covariance of X with S could be computed as

$$\text{cov}(X, S) = \frac{\sum_{i=1}^5 X_i \left(\sum_{j=1}^5 \sum_{k=1}^5 S_{ijk} \right)}{25} \quad (42)$$

where i denotes the row, j denotes the column, and k denotes the individual value of S within each cell. However, since the sum of Z is 0, $\sum_{k=1}^5 S_{ijk}$ is simply five times the value in the corresponding cell of table 4. Thus

$$\text{cov}(X, S) = \sum_{i=1}^5 X_i \left(\sum_{j=1}^5 5 \bar{S}_{ij} \right) / 25 \quad (43)$$

$$= \sum_{i=1}^5 X_i \left(\sum_{j=1}^5 \bar{S}_{ij} \right) / 5 \quad (44)$$

where \bar{S}_{ij} represents the value in the ith row and jth column of table 4.

The variance of the values in the cells of table 4 is 1.05. This figure can be computed directly from the

cells of table 4 or by adding the variances of X and Y given on p. 40. Were all of the values of S_{ijk} used to calculate the variance of S, the result would be

$$\sigma_s^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = 0.79 + 0.26 + 0.76 = 1.81 \quad (45)$$

The correlation of X and S, ρ_{XS} , is $\text{cov}(X,S)/\sigma_x \cdot \sigma_s$ and is, in this case, 0.66, the expected value. Thus, the information in table 4 suffices to determine the correlations provided the variance due to Z can be determined.

The values in table 4 also suffice to represent expected scores, as the mean of the individual values of S_{ijk} in a cell is S_{ij} .

A linear transformation applied to the cells of table 4 gives them the appearance of expected scores for an item scored as either 0 (incorrect) or 1 (correct). Applying equation (38) to the cells of table 4 yields table 5. The marginal values of table 4 are divided by their respective standard deviations to give z-scores.

		Y				
		-1.46	-0.60	0.00	0.60	1.46
X	-1.46	0.35	0.38	0.40	0.43	0.46
	-0.60	0.40	0.44	0.46	0.48	0.52
	0.00	0.44	0.48	0.50	0.52	0.56
	0.60	0.48	0.52	0.54	0.56	0.60
	1.46	0.54	0.57	0.60	0.62	0.65

table 5

Table 5 can be interpreted as follows: the marginal values, labeled X and Y represent an individual's standing with respect to these two traits; the numbers given are z-scores. The values in the cells of the table represent levels of difficulty for persons as they are classified by the variables X and Y. For example, persons who are 0.60 standard deviations above the mean with respect to trait X and 1.46 standard deviations below the mean with respect to trait Y have a probability of 0.48 of giving a "correct" response to the item in question.

Table 5 completes the purpose of this section: it demonstrates that item characteristics defined by a vector model can be translated to statistics commonly used to describe test items.

ITEM SELECTION: INTERNAL CRITERION

A person's expected score on a test composed of all 225 items is simply the sum of his expected scores on the individual items. The sum of the 225 expected score matrices for the items is an expected score matrix for the whole test. The covariance of an individual item with the whole test can be found by multiplying the values in the cell of the item matrix by the corresponding cell of the whole-test matrix, adding these values, dividing by 225 and subtracting the product of the respective mean scores. That is,

$$\text{cov}(\text{item, whole test}) =$$

$$\sum_{i=1}^{15} \sum_{j=1}^{15} x_{ij} X_{ij} / 225 - \bar{x}_{..} \bar{X}_{..} \quad (46)$$

where x_{ij} represents the expected score on an item and X_{ij} represents the expected score on the whole test for persons of the same characteristic.

The usual procedure of item analysis using an internal criterion is to select those items whose correlation coefficient with the whole test is greatest. The procedure in this demonstration is to select those items for which the covariance of the expected scores is greatest. Two differences are evident: (1) the present procedure uses covariances rather than correlations and (2) the

covariance is computed in terms of expected scores rather than actual scores. That these changes do not affect the selection procedure is easily demonstrated.

The correlation of two variables, say X and Y, is simply the covariance divided by the product of the standard deviations of X and Y. That is

$$\rho_{XY} = \text{cov}(X, Y) / \sigma_X \cdot \sigma_Y \quad (47)$$

In the present case, σ_X , the standard deviation for the item, is the same for all items being compared because all item matrices have been "standardized" to have a mean of 0.5 and a standard deviation of 0.1 (p. 40). The standard deviation of the whole test, σ_Y , is the same in all cases because the same whole-test matrix was used in all cases. Thus, the correlation is directly proportional to the covariance; the order of the items is the same in terms of correlation or covariance.

The relationship of the covariance of expected scores and the covariance of the actual scores is more easily treated if each observed score, X_0 or Y_0 , is considered to be the sum of two components: an expected score (or true score), X_t or Y_t , and an error term, e_X or e_Y .

Thus

$$X_0 = X_t + e_X, \quad (48)$$

$$Y_0 = Y_t + e_Y. \quad (49)$$

No generality is lost if X and Y are chosen so that

$$\bar{X} = \bar{Y} = 0 \quad (50)$$

The covariance of the observed scores is given by

$$\text{cov}(X_o, Y_o) = \frac{1}{n} \sum (X_o \cdot Y_o) \quad (51)$$

$$= \frac{1}{n} \sum (X_t + e_x)(Y_t + e_y) \quad (52)$$

$$= \frac{1}{n} \sum (X_t Y_t + X_t e_y + e_x Y_t + e_x e_y) \quad (53)$$

Generally, it might be assumed that e_x and e_y are normally distributed with means of 0 and that they are independent of one another and of the true scores. In that case, the last three terms of (53) disappear. In the present case, e_x and e_y are not random variables, but are rigidly symmetric with means of 0. Implicit in the summation sign of (53) is the summing over all values of e_y for each value of X_t in the second term. As $\sum e_y = 0$ ($\sum Z = 0$, p. 29) the second term of (53) disappears. Similarly, the third and fourth terms disappear. Thus

$$\text{cov}(X_o, Y_o) = \text{cov}(X_t, Y_t) \quad (54)$$

That is, the covariances obtained using expected scores are the same as the covariances that would have been obtained using actual scores.

The hundred items having the greatest covariance with the whole test are selected. In cases where the covariance of the 100th and the 101st items have equal covariances, the 101st item is also selected. No case

is encountered in this study where the 100th, 101st, and 102nd items have equal covariance.

ITEM SELECTION: EXTERNAL CRITERIA

Items are also ranked in terms of a weighted sum of their covariance with the job elements. The covariance of each item matrix with each of the two elements is calculated from the expected score matrix of the item (like table 5). The covariances are weighted according to the relative importance of the elements. For example, if factor 1 is deemed twice as important as factor 2, the covariance of the item with factor 1 is multiplied by two-thirds and the covariance with factor 2 by one-third. The weighted covariances are added to provide a "synthetic covariance." About 100 items are selected on the basis of the synthetic covariance just as they are for covariance with the whole test. As with the whole test covariance, if the 100th and 101st items have equal values, the 101st item is included.

This procedure will be referred to as "synthetic item analysis" or simply "synthetic analysis" in the remainder of this paper.

CONSTRUCTION OF A VALIDATION CRITERION

The construction of a matrix to use as a criterion for validation uses earlier assumptions regarding the factor structure of job performance: that job performance depends upon factors 1 and 2, that the relationship of these factors is as specified by the vector model regarding their correlation and relative importance and that there are no systematic sources of error.

The validation matrix is constructed to represent the vector sum of the factor-vectors. The correlation of the vector sum to factor 1, ρ_{XY} , is simply the cosine of the angle between these vectors. A second correlation, ρ_{YV} , is taken to be the sine of this angle. A matrix of "expected scores" is constructed by precisely the same procedure that was described on pp. 36-40, above. The value of ρ_{XS} in that procedure is replaced by ρ_{XV} , ρ_{YS} is replaced by ρ_{YV} , and ρ_{XY} is taken to be 0. In the demonstration, the identical Fortran subroutine that is used to create item matrices is used to create the validation matrix.

VALIDATION

An expected score matrix for each set of one hundred (or 101) items is constructed by adding their respective item matrices. This process is identical to the procedure used to develop an expected score matrix for all 225 items.

The covariance of each of these matrices with the validation matrix is calculated as

$$\text{cov}(X_T, X_C) = \frac{1}{225} \sum_{i=1}^{15} \sum_{j=1}^{15} (X_{T_{ij}} X_{C_{ij}}) - 0.25 n \quad (55)$$

where $X_{T_{ij}}$ represents an expected score on the test, $X_{C_{ij}}$ represents a corresponding value from the validation matrix, and n is the number of items on the test. The last term, $0.25n$, is the product of the means of the elements of both matrices. Each item matrix has a mean of 0.5, as does the validation matrix. The test expected score matrix is the sum of n item matrices and therefore has a mean of $0.5n$. The product of this and the mean of the validation matrix is $0.25n$.

The standard deviation of the test expected score matrix can be readily calculated using the values in that matrix. However, this value is not the standard deviation of actual scores; variance due to error (Z) is omitted. To correct this, that part of the total variance of each

item matrix is determined according to the following argument.

The item matrices were initially constructed by determining three variables, X, Y, and Z, such that the correlations of X and Y with the sum, S, and the correlation of X with Y had certain proscribed values. The correlations of X and Y with Z were to be 0. The variance of S is given by

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + 2\sigma_X\sigma_Y\rho_{XY} \quad (15)$$

That part of the variance due to Z, then, is σ_Z^2/σ_S^2 . As the matrix is transformed to have a total standard deviation of 0.1, the variance of the transformed matrix, including any variance due to Z, becomes 0.01 and the variance of each item matrix due to Z is $0.01 \sigma_Z^2/\sigma_S^2$.

It is assumed that the errors for different items are independent. Therefore, the variance of the sum of n items due to error is $0.01 \cdot \sum_{i=1}^n \sigma_{Z_i}^2/\sigma_{S_i}^2$. The variance of the actual scores is given by

$$\sigma_A^2 = \sigma_E^2 + 0.01 \sum_{i=1}^n \sigma_{Z_i}^2/\sigma_{S_i}^2. \quad (56)$$

where σ_A^2 is the variance of the actual rather than expected scores and σ_E^2 is the variance of expected scores as calculated from the values in the expected score matrix.

As the standard deviation of the criterion matrix is 0.1, the validity coefficient of either test is given by

$$\rho_{TC} = \text{cov}(X_T, X_C) / 0.1 \sigma_A \quad (57)$$

where $\text{cov}(X_T, X_C)$ is given by equation (55) and σ_A is the square root of σ_A^2 given in (56).

In order for the validity coefficients to be comparable, they have to be based on tests of the same size. Thus, ρ_{TC} must be corrected if it is based on a test of 101 items rather than 100 items. This is done by the Spearman-Brown prophecy formula (6) where $n = 1/1.01$. That is

$$\rho'_{TC} = \frac{(\frac{1}{1.01})\rho_{TC}}{1 + (\frac{1}{1.01} - 1)\rho_{TC}} \quad (58)$$

where ρ'_{TC} is the corrected value of ρ_{TC} .

For both tests, ρ_{TC} (or ρ'_{TC} where $n = 101$) is calculated for the values of the parameters as indicated on pp. 33-34.

The Fortran program used to perform this procedure is included as an appendix.

CHAPTER IV

RESULTS

The values of the "whole-test reliability," the consequent "typical angle," and the resulting validity coefficients for a test of one hundred items chosen by an internal criterion and by synthetic analysis are given in table 6.

Reli- ability	Angle	Validity: internal criterion	Validity: synthetic analysis
.99375	52.1 ^o	.852	.940
.9875	63.8	.851	.946
.975	73.7	.849	.948
.95	80.8	.847	.948
.90	85.2	.846	.944
.80	87.8	.846	.940
.60	89.1	.846	.940

table 6

All of the measures indicated in table 6 are made with the mean of the item-vectors aligned with factor-vector 1, with an angle of 90^o between the factor-vectors, and with factor-vector 1 twice as long as factor-vector 2.

The validities of table 6 are presented graphically in figure 5. It should be noted that equal distances on the horizontal scale of figure 5 do not represent equal intervals in terms of reliability or angle. The reader

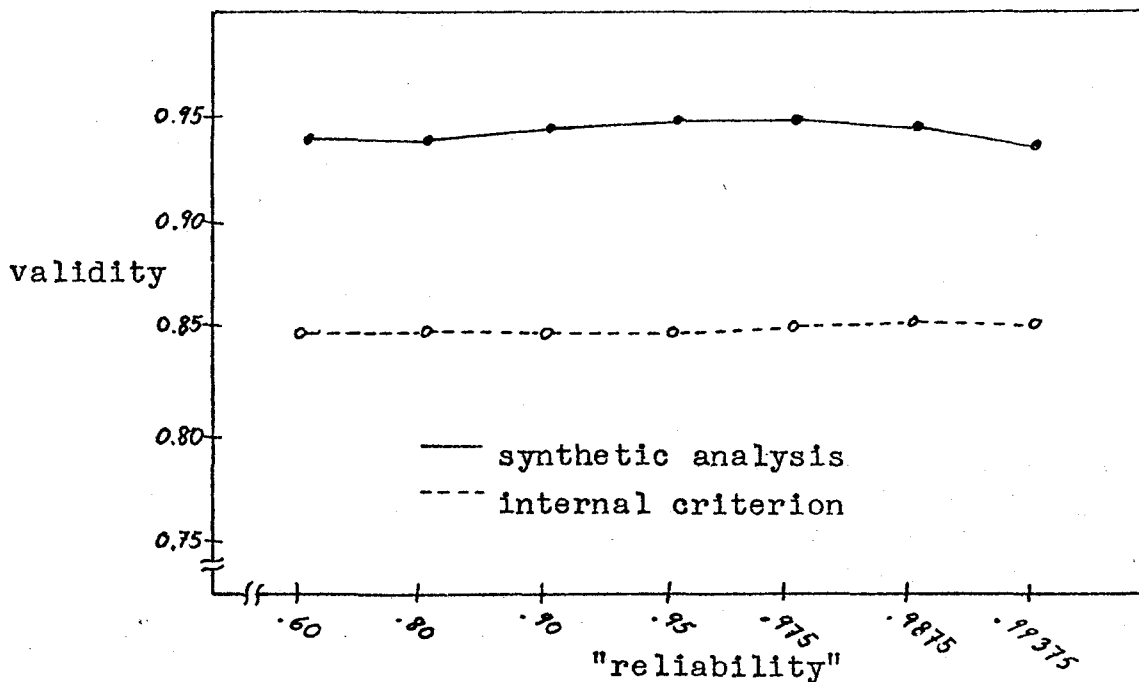


figure 5

should also note that the vertical scale does not start at zero.

The relative effectiveness of the two methods can also be evaluated by comparing resulting "signal-to-noise" ratios. Usually, a signal-to-noise ratio is the variance due to true score variance divided by variance due to error; it is usually a measure of reliability. The present case requires a modification of the definition to make it applicable to validity. For the present usage, the signal-to-noise ratio will be defined as σ_r^2/σ_e^2 where σ_r^2 is the variance of the observed scores due to variance of the trait(s) that the test is intended to measure and

σ_e^2 is variance due to both systematic and unsystematic error.

Applying this statistic to data from table 6, for a whole-test "reliability" of 0.90, the signal-to-noise ratio for the test created by an internal criterion is 2.94. For the test created by synthetic analysis, it is 8.19.

All of the item-vectors are symmetrically arranged about some central vector, referred to in this paper as the "overall direction." As defined on p. 33, this vector is always in the same plane as the factor-vectors. Its measure indicates the degrees of rotation from factor 1 toward factor 2. The values used in this demonstration, and the consequent validities, are given in table 7.

Overall direction	Validity: internal criterion	Validity: synthetic analysis
-45°	.328	.936
-30	.547	.930
-15	.707	.941
0	.846	.944
15	.928	.949
30	.948	.948
45	.909	.945

table 7

All of the measures in table 7 are made with a "whole-test reliability" of 0.90, with the factor-vectors perpendicular to one another, and with factor-vector 1 twice as long as factor-vector 2.

The validities of table 7 are presented graphically in figure 6. In addition to the angle of the overall direction to factor-vector 1, the absolute value of the angle between the overall direction and the vector sum of the factor-vectors is given. The latter figures are in parentheses.

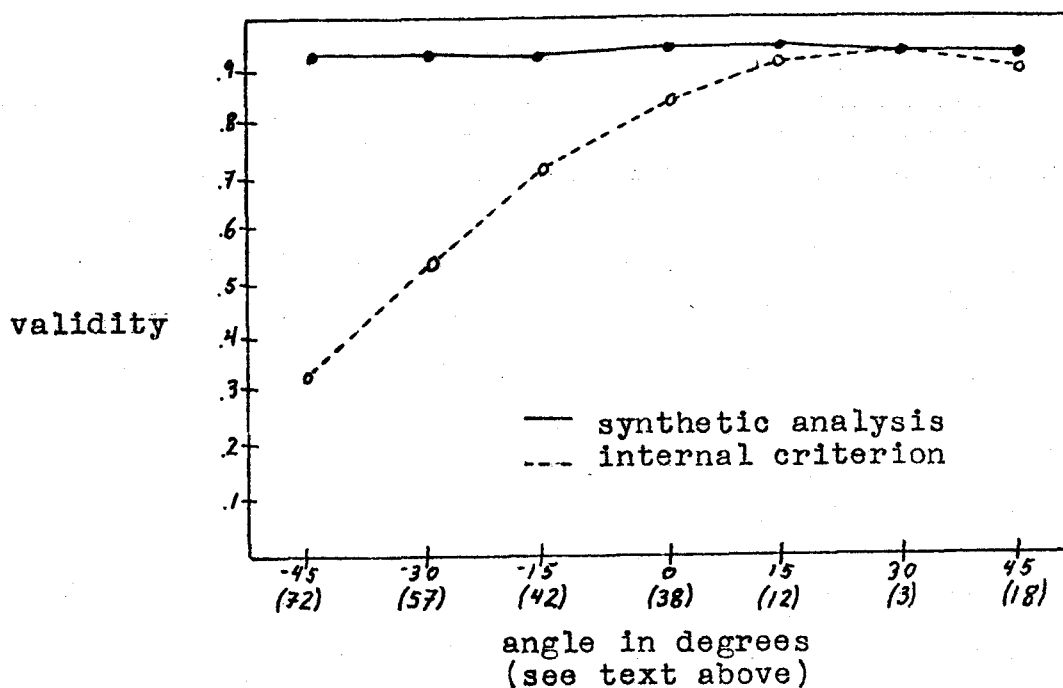


figure 6

Table 8 gives the values assigned to the length of factor-vector 1, with consequent validities. As factor-vector 2 has a length of one throughout the demonstration, the length of factor-vector 1 is in fact the ratio of the length of factor-vector 1 to that of factor-vector 2 or, equivalently, the relative importance of factor 1 to factor 2.

Relative importance 1 to 2	Validity: internal criterion	Validity: synthetic analysis
0.25	.227	.929
0.50	.421	.914
1.00	.668	.935
2.00	.846	.944
4.00	.919	.947

table 8

All of the measures indicated in table 8 are made with a "reliability" of 0.90, the overall direction of the item-vectors aligned with factor-vector 1, and with the factor-vectors perpendicular to one another.

The data in table 8 are presented graphically in figure 7. It should be noted that equal distances along the horizontal axis do not represent equal intervals in the independent variable.

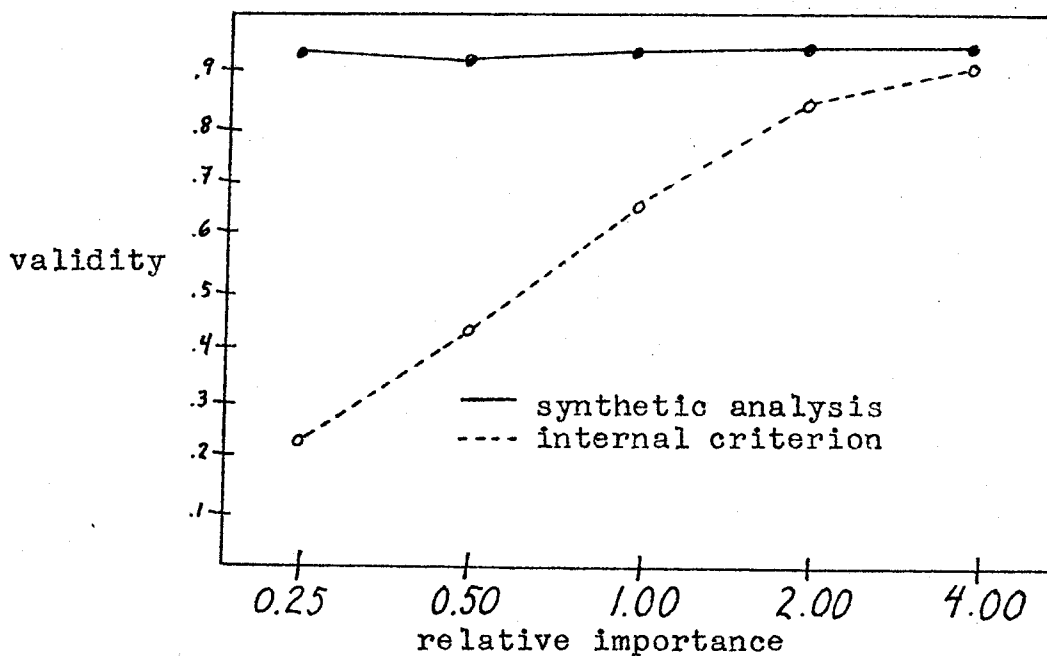


figure 7

The effect of changing the angle between the factor-vectors is examined under two conditions:

1) with the size of factor-vector 1 twice that of factor-vector 2, the overall direction of the item-vectors aligned with factor 1, and a "reliability" of 0.90, and

2) with the size of factor-vector 1 twice that of factor-vector 2, the overall direction of the item-vectors aligned with factor 1, and a "reliability" of 0.99375.

Table 9 gives the angles between the two factor-vectors and the consequent validities for the first set of conditions. Table 10 gives the angles and consequent validities for the second.

Inter-factor angle	Validity: internal criterion	Validity: synthetic analysis
90°	.846	.944
80	.836	.942
70	.917	.947
60	.949	.936
50	.946	.922
40	.854	.875

table 9

Inter factor angle	Validity: internal criterion	Validity: synthetic analysis
90°	.852	.940
80	.865	.950
70	.919	.953
60	.953	.937
50	.931	.931
40	.900	.911

table 10

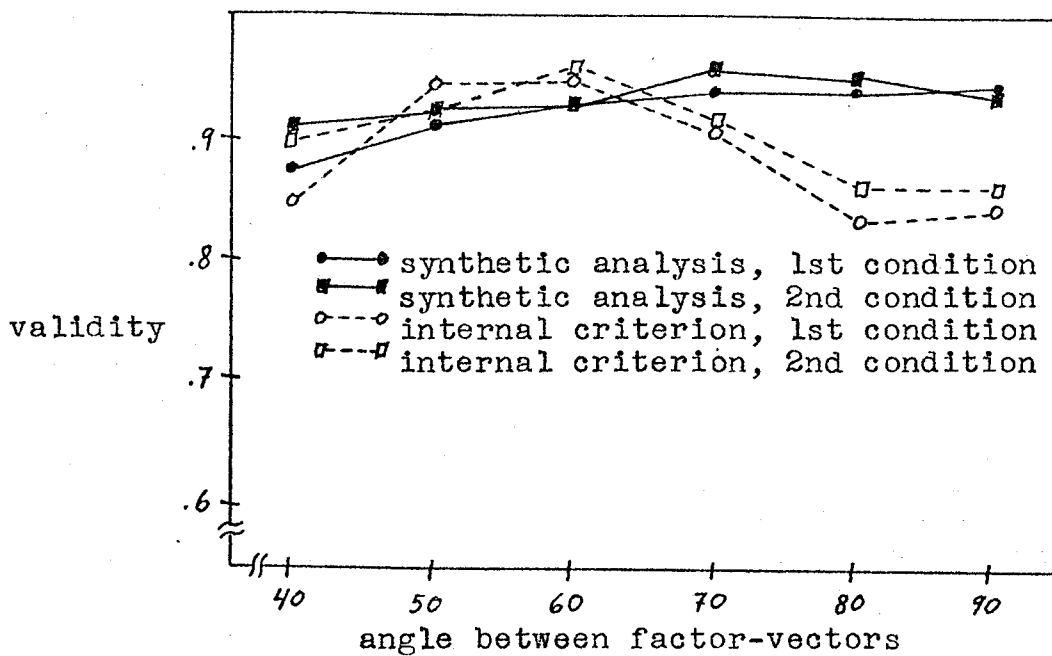


figure 8

The data from tables 9 and 10 are presented graphically in figure 8.

CHAPTER V

DISCUSSION

INTERPRETATION OF RESULTS

Generally, data from experiments does not exactly match predicted or expected values. Deviations from an expected or mean value are attributed to "random error," the effect of uncontrolled variables. Deviations from expected values in the present data cannot be so excused, as there is no "random" error. The computer hardware, software, and program completely determine the outcome.

There are, however, some effects that are peculiar to a particular configuration of hardware, software, and program which have no general significance. A particular validity coefficient for a set of a hundred items, for example, depends upon the exact angle between the criterion vector and the item-vectors. A shift of 2° in the relative location of the criterion and item-vectors would change the validity coefficient slightly (by a factor of about 0.999). This large a shift would probably not be great enough to affect the set of items selected, nor would it have any practical significance. These effects, peculiar to the immediate situation and of no general

interest, may be thought of as "error" even though they are not random in the usual sense.

An inspection of validity coefficients for the tests selected by synthetic analysis in tables 6 and 7 suggest that this "error" can be as large as 0.01 when measured from an average value. Tables 8, 9, and 10 suggest a somewhat larger value.

The effect of changing the average reliability or spread of the items is shown in table 6 (figure 5). This table indicates that over the range of item spread which might reasonably be expected and under the circumstances given, the validity of a test selected by synthetic analysis is appreciably greater than that of a test selected by an internal criterion. It also indicates that there are only minor drops in validity for a test selected by an internal criterion with large drops in reliability. For a test selected by synthetic analysis, these data do not indicate any relationship between validity and item spread.

Logically, changing the arrangement of the items affects tests selected by synthetic analysis only by limiting the number and symmetry of items grouped around the criterion. That is, a test produced by synthetic analysis will have optimal validity (1) if there are an ample number of items which are highly correlated with

the criterion and (2) if these items are symmetrically arranged about the criterion so that their individual biases will add to zero. Inspection of figure 3 (p. 29) suggests that rotating the items clockwise moves away from this optimum. Rotating toward the factor-vector sum (the validation criterion) should produce optimal validity. The validities for synthetic analysis in table 7 are consistent with this argument, but indicate that its effect is slight under the conditions given.

Because the set of items selected by an internal criterion is not responsive to the location of job factors, the validity of such a test decreases as the angle between the overall direction and the vector representing the sum of the factor-vectors increases. Table 11 shows the overall direction, as indicated in table 7, the angle in degrees between the overall direction and the validation criterion, ϕ , the cosine of this angle, and the ratio of the validity of a test selected by an internal criterion to its maximum value. These figures suggest that the

Overall direction	ϕ	$\cos(\phi)$	Ratio V/V_{\max}
-45°	72°	0.31	0.35
-30	57	0.54	0.58
-15	42	0.74	0.75
0	27	0.89	0.89
15	12	0.98	0.98
30	3	1.00	1.00
45	18	0.95	0.96

table 11

validity of a test is proportional to the cosine of an angle between the vector sum of its items and the vector sum of the job factors. These data also suggest that, while the validity of a test selected by an internal criterion is limited by its reliability, the validity of the initial set of items, taken as a whole, can be a much more important consideration.

Changing the relative importance of the factors has no apparent effect upon the validity of a test selected by synthetic analysis, as indicated by table 8 (figure 7).

For a test selected by an internal criterion, however, the validity increases as the ratio of factor 1 to factor 2 increases. It is important to recall that the overall direction of the items (the internal criterion) is aligned with factor 1. It is noted above that the validity of a test selected by an internal criterion is proportional to the cosine of the angle between the internal criterion and the validation criterion. Changing the relative size of the factor-vectors changes the angle of their sum (the validation criterion) with factor-vector 1 (the internal criterion). Table 12 gives the ratio of factor 1 to factor 2, as indicated in table 8, the consequent angle of the validation criterion to the internal criterion, θ , and the corresponding validity divided by 0.948, the maximum validity in table 7. A comparison of the data in table

Factor 1/ Factor 2	ϕ	$\cos(\phi)$	Ratio: V/V_{\max}
0.25	76°	0.24	0.24
0.50	63	0.45	0.44
1.00	45	0.71	0.70
2.00	27	0.89	0.89
4.00	14	0.97	0.97

table 12

12 with that in table 11 supports the hypothesis that the validity of a test chosen by an internal criterion is proportional to the angle between the overall direction and the validation criterion.

Tables 9 and 10 indicate that the validity of a test selected by synthetic analysis declines slightly as the angle between the factor-vectors decreases. This phenomenon is not justified on a theoretical basis. In terms of the vector model, the function used to select items is

$$f(\theta, \phi) = \frac{2}{3} \cos \theta + \frac{1}{3} \cos \phi \quad (59)$$

where θ is the angle between the item-vector and factor-vector 1 and ϕ is the angle between the item-vector and factor-vector 2. It is easy to demonstrate with a few hypothetical values that this function has a maximum for values of θ and ϕ which correspond to the sum of the factor-vectors (the validation criterion). Moreover, this function is symmetric about its maximum. These are precisely the characteristics desired in a selector function:

it has a maximum at the appropriate point and its value decreases with increased distance from the maximum, regardless of direction.

Apart from any failure of synthetic analysis to select an optimal set of items, there is reason to expect validity to decrease with the angle between the factor-vectors. This model is constructed by adding three random variables, X, Y, and Z, where X represents the effect of factor 1, Y the effect of factor 2, and Z, error. As the X and Y addends become more highly correlated, their effects "overlap" and the sum of their effects contributes proportionally less to the total variance. As the variance of each item is arbitrarily set at 0.01, this means that the role of error is increased on an absolute as well as a relative basis. Increasing the role of error lowers reliability and therefore lowers optimum validity.

The mechanics of lowering the relative contribution of X and Y can perhaps be better seen in equation (35):

$$\sigma_x^2 = \rho_{xS}^2 \sigma_z^2 / (1 + A\rho_{xy} + A^2\rho_{xy}^2) - \rho_{xS}^2 (1 + 2A\rho_{xy} + A^2) \quad (35)$$

The value of σ_x^2 decreases as the denominator of (35) increases. That denominator increases as ρ_{xy} increases, for all values of A and ρ_{xS}^2 applicable to the items selected. As σ_y^2 is a linear function of σ_x^2 , its magnitude also decreases as ρ_{xy} increases. Since σ_z^2 is constant for items not in the factor-vector plane, its relative magnitude increases as the magnitude of σ_x^2 and σ_y^2 decreases. Though

the mathematical proof of these comments is straightforward, it is lengthy.

Thus, the decline in the validity of a test selected by synthetic analysis as the angle between the job factors decreases appears to be an artifact of this particular model. In practical applications, it does not seem likely that the magnitude of random error would be a function of the intercorrelation of job factors identified. If it were, of course, the model would be most appropriate.

There are two mechanisms which account for the relationship of the validity of a test selected by an internal criterion and the angle between the job factors. First, rotating factor-vector 2 toward factor-vector 1 has the effect of rotating the sum of those vectors, the validation criterion, toward the overall direction of the items. As noted above, the smaller the angle between the validation criterion and the overall direction of the items, the greater the validity. Second, increasing the correlation of the factors has the effect of weighting the factors: if the angle of factor-vector 2 to factor-vector 1 is 60° , for example, a unit gain in the direction of factor-vector 2 represents a half-unit gain in the direction of factor-vector 1. Thus, for the angle mentioned, factor 2 is weighted half as much as factor 1. This is wholly consistent with the hypothesis that factor 2 is half as important as factor 1. For angles other than 60° , selection

by an internal criterion is less than optimal. The data presented on tables 9 and 10 (figure 8) are consistent with both of these mechanisms.

A problem inherent in the use of any external criterion is that that criterion may have limited validity. This can be shown in terms of a Venn diagram (figure 9).

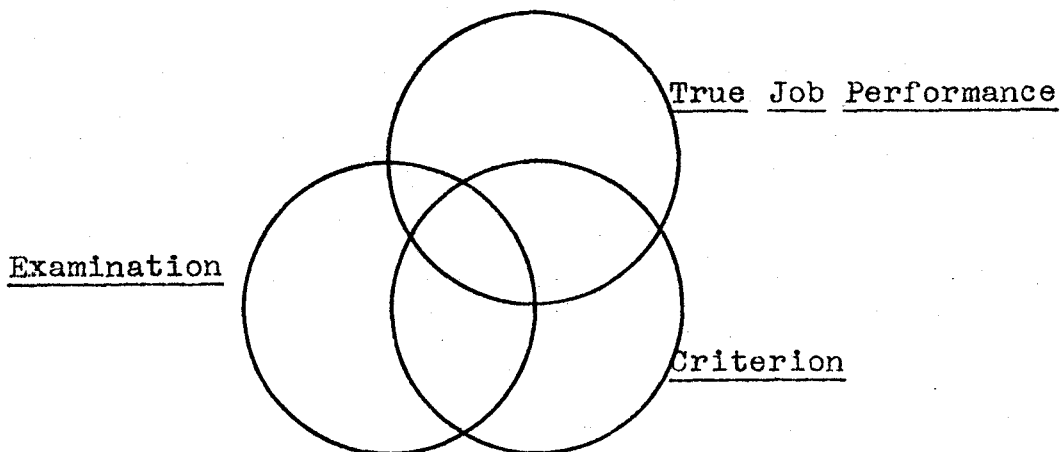


figure 9

In this diagram, the overlap of the two circles represents the correlation of the corresponding measures.

If there were a perfect measure of true job performance, the result of that measure should correlate positively with the criterion score. Unfortunately, no such measure exists and the magnitude of that correlation can only be estimated on the basis of content analysis. The correlation of the examination score and the criterion score is measured empirically but is of no interest in itself. What is actually wanted is the correlation of the examination score and true job performance. As seen

in figure 9, the correlation of the examination score with the true job performance is only partially determined by the correlation of the examination with the criterion. Having no way to determine the relationship of true job performance to the criterion or to the examination, one can only assume, optimistically, that the criterion is very nearly the same as true job performance. Estimates of the criterion validity of an examination reflect this optimism, as does the synthetic analysis procedure described in this paper.

Inherent in the synthetic analysis procedure is the development of multiple criteria. This procedure may require a more analytic consideration of job requirements than the selection of a single criterion and thereby improve the quality of the criteria. Nonetheless, the quality of a test selected by synthetic analysis is dependent upon the quality of the criterion measures of the job factors. All conclusions drawn from this study must be tempered by this consideration.

APPLICATION

The mechanisms of applying synthetic analysis to a practical situation are straightforward: To develop a selection instrument for a job, it is first necessary to determine the factors or personal traits which characterize successful workers. Some criteria must be developed to determine the degree to which a worker has these characteristics. This will generally consist of some sort of supervisory rating. Items are developed and tried on a sample of workers or prospective workers. The items are correlated against criterion scores, weighted according to the judged importance of the factors, and added to yield a synthetic correlation. Items are chosen to be included in a final version of the test according to their synthetic correlation.

A problem with this procedure is that it seems unnecessarily complicated. The same result could be achieved, theoretically, by determining one criterion for job success, thus eliminating the need to determine and weight job factors and to correlate items against several criteria rather than one criterion.

The advantage of using synthetic analysis and job factors is clearer in situations where selection devices need to be made for several jobs with overlapping factors.

If these examinations are developed sequentially, each subsequent effort can borrow from preceeding studies. If criteria and items have been developed to measure conscientiousness, for example, in the first examination, the same criteria and items can be adapted to later situations.

It may be more advantageous, however, to develop selection devices concurrently rather than sequentially. A group of jobs may be identified which are expected to overlap in terms of required worker characteristics. Criteria developed to measure a factor will hopefully function for all of the jobs for which the factor is pertinent. Working with several jobs at once may help to make the factors and items less job specific. Furthermore, working with several jobs at once provides larger samples of workers upon which to try items. These samples may provide a wider range of a factor, thus further increasing the precision of correlation procedures.

Where several jobs are investigated at the same time, it may be feasible to develop one examination with several scoring keys. From a set of, say, 100 items given to all applicants, there may be 70 which synthetic analysis identifies as a test for job A, a different but overlapping set of 65 items which are identified as a test for job B, and so forth. Any applicant could readily be given scores for all of the jobs covered by the basic set

of 100 items. This procedure is similar in principle to criterion keying procedures used in various personality and interest inventories.³⁴

The preceding discussion of applications can be made to fit educational problems by merely changing the terminology. Whether a person is applying for a job or is being considered for a reading program or graduate study, the statistical procedures involved in forecasting success are the same. In the area of graduate study, for example, different characteristics of successful students could be identified by experienced teachers, administrators, and students. Undoubtedly, there are some characteristics which are factors of success in any discipline. Tenacity, for example, might be a major factor in determining the success of a doctoral candidate whether he studies astronomy or ancient history. It is equally certain that some factors are more important to some disciplines than others. For example, the ability to read and remember large volumes of literature may be more important to a historian than to a physicist.

A graduate school selecting doctoral candidates is in the position of an employer selecting workers. There are several programs into which a candidate may enter just

³⁴Anne Anastasi, Psychological Testing (New York: The Macmillan Company, 1968), p. 440 ff.

as an employer may have several kinds of jobs to be done. For groups of programs requiring similar characteristics, giving all of the candidates the same items and scoring the items shown to be measures of potential in a particular program seems a reasonable strategy. Having the capacity to differentially forecast success in various programs should be a great benefit to both the student and the educator.

BIBLIOGRAPHY

- Anastasi, Anne. Psychological Testing. New York: The Macmillan Company, 1968.
- Balma, Michael J. "The Concept of Synthetic Validity." Personnel Psychology, XII (Autumn, 1959), 395-96.
- Bass, Bernard M. "Further Evidence on the Dynamic Character of Criteria." Personnel Psychology, XV (Spring, 1962), 93-97.
- Cronbach, Lee J. and Warrington, Willard G. "Efficiency of Multiple-Choice Tests as a Function of Spread of Item Difficulties." Psychometrika, XVII (June, 1952), 127-47.
- Ebel, Robert L. Essentials of Educational Measurement. Englewood Cliffs: Prentice-Hall, Inc., 1972.
- Englehart, Max D. "A Comparison of Several Item Discrimination Indices." Journal of Educational Measurement, II (June, 1965), 69-76.
- Fossum, John A. "An Application of Techniques to Shorten Tests and Increase Validity." Journal of Applied Psychology, LVII (February, 1973), 90-92.
- Ghiselli, Edwin E. "The Generalization of Validity." Personnel Psychology, XII (Autumn, 1959), 397-402.
- Guion, Robert M. "Synthetic Validity in a Small Company: A Demonstration." Personnel Psychology, XVIII (Spring, 1965), 49-63.
- Hasson, David J. "An Evaluation of Two Methods of Test Item Selection." Dissertation Abstracts, Vol. 32A 6200-A.
- Henrysson, Sten. "Gathering, Analyzing, and Using Data on Test Items." Educational Measurement. Edited by Robert L. Thorndike. Washington, D. C.: American Council on Education, 1971.

- Kirchner, Wayne K. and Reisberg, Donald J. "Differences between Better and Less-Effective Supervisors in Appraisal of Subordinates." Personnel Psychology, XV (Autumn, 1962), 295-302.
- Lawshe, C. H. and Steinberg, Martin D. "Studies in Synthetic Validity. I. An Exploratory Investigation of Clerical Jobs." Personnel Psychology, VIII (1955), 291-301.
- Lord, Frederic M. "The Relation of the Reliability of Multiple-Choice Tests to the Distribution of Item Difficulties." Psychometrika, XVII (June, 1952), 181-94.
- Mosier, Charles I. "A Note on Item Analysis and the Criterion of Internal Consistency." Psychometrika, I (December, 1936), 275-82.
- Nunnally, Jum C. Educational Measurement and Evaluation. New York: McGraw-Hill Book Company, 1972.
- Primoff, Ernest S. "Basic Formulae for the J-coefficient to Select Tests by Job Analysis Requirements." Washington, D. C.: Test Development Section, United States Civil Service Commission, 1955.
- _____. "The J-coefficient Approach to Jobs and Tests." Personnel Administration, XX (May-June, 1957), 34-40.
- Richardson, M. W. "Notes on the Rationale of Item Analysis," Psychometrika, I (1936), 69-76.
- _____. "The Relation between the Difficulty and the Differential Validity of a Test." Psychometrika, I (June, 1936), 33-49.
- Ryans, David G. "The Results of Internal Consistency and External Validation Procedures Applied in the Analysis of Test Items Measuring Professional Information." Educational and Psychological Measurement (1951), 549-560.
- Selby, Dane. The Validation of Tests Using J-Coefficient: A Feasibility Study. Illinois: Research and Test Development, Illinois Department of Personnel, 1975.
- Stanley, Julian C. "Reliability." Educational Measurement. Edited by Robert L. Thorndike. Washington, D. C.: American Council on Education, 1971.

APPENDIX

The following is the Fortran program used to generate and evaluate hypothetical data as described in chapter III.

Because up to 72 columns can be used on a Fortran card and only about 60 columns may be typed on these pages, the arrangement of continuation cards has been altered in some cases. A "&" in the sixth column indicates a continuation of the previous line. "Ø" represents the number zero.

At the end of the program is a glossary of Fortran variables used in this program.

```
DIMENSION Z (15),ZITMFP(15), XITMF1(15),XITMF2(15),  
&RSX(15,15),RSY(15,15),EXSCOR(15,15),COVEX(15,15),  
&COVSYN(15,15),IM(225),JM(225)
```

```
Z(1) = -1.83  
Z(2) = -1.28  
Z(3) = -Ø.97  
Z(4) = -Ø.73  
Z(5) = -Ø.52  
Z(6) = -Ø.34  
Z(7) = -Ø.17  
Z(8) = Ø.ØØ  
Z(9) = Ø.17  
Z(1Ø) = Ø.34  
Z(11) = Ø.52  
Z(12) = Ø.73  
Z(13) = Ø.97  
Z(14) = 1.28  
Z(15) = 1.83
```

```
VZ = Ø.9168
```

```

98 READ (5,110) RTEST,TITMF1,ALF1F2,SIZE1
110 FORMAT (5X,4F9.7)
   IF(RTEST) 99,99,97
97 CONTINUE

   SIZE2 = 1.0
   RITEM = RTEST/(100.0 - 99.0 * RTEST)
   TYPAL = ARCOS(RITEM)

   DO 1 I = 1,15
1  ZITMFP(I) = TYPAL * Z(I)
   DO 2 I = 1,15
2  XITMF1(I) = ZITMFP(I) + TITMF1(I)
   DO 3 I = 1,15
3  XITMF2(I) = ALF1F2 - XITMF1(I)

   DO 4 I = 1,15
   DO 5 J = 1,15
5  RSX(I,J) = COS(ZITMFP(J))*COS(XITMF1(I))
4  RSY(I,J) = COS(ZITMFP(J))*COS(XITMF2(I))
4  CONTINUE

   RXY = COS(ALF1F2)

   DO 6 I = 1,15
   DO 7 J = 1,15
7  EXSCOR(I,J) = 0.0
6  CONTINUE

   DO 10 I = 1,15
   DO 11 J = 1,15
   RX = RSX(I,J)
   RY = RSY(I,J)
   CALL DEVIAT (RX,RY,RSY,VZ,DX,DY,SDS,PVZ)
   DO 18 K = 1,15
   DO 19 L = 1,15
   A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5
   EXSCOR(K,L) = EXSCOR(K,L) + A
19 CONTINUE
18 CONTINUE
11 CONTINUE
10 CONTINUE

   DO 22 I = 1,15
   DO 23 J = 1,15
   RX = RSX(I,J)
   RY = RSY(I,J)
   CALL DEVIAT (RX,RY,RXY,VZ,DX,DY,SDS,PVZ)
   COVEX(I,J) = 0.0
   COVSYN(I,J) = 0.0

```

```

DO 30 K = 1,15
DO 31 L = 1,15
A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5
COVEX(I,J) = COVEX(I,J) + A*EXSCOR(K,L)
COVSYN(I,J) = COVSYN(I,J) + SIZF1/(SIZF1 + SIZF2)*
&Z(K)*A + SIZF2/(SIZF1 + SIZF2)*Z(L)*A
31 CONTINUE
30 CONTINUE

COVEX(I,J) = COVEX(I,J)/225.0 - 56.25
COVSYN(I,J) = COVSYN(I,J)/225.0
23 CONTINUE
22 CONTINUE

DO 40 M = 1,100
XMAX = 500.0
DO 35 I = 1,15
DO 36 J = 1,15
IF(XMAX - COVEX(I,J)) 37,36,36
37 XMAX = COVEX(I,J)
XMOX = XMAX
IM(M) = I
JM(M) = J
36 CONTINUE
35 CONTINUE
I = IM(M)
J = JM(M)
COVEX(I,J) = -500.0
40 CONTINUE

M = 101
DO 38 I = 1,15
DO 39 J = 1,15
IF(XMOX - COVEX(I,J)) 41,41,39
41 IM(M) = I
JM(M) = J
M = M + 1
39 CONTINUE
38 CONTINUE
M = M - 1
XM = M

DO 42 K = 1,15
DO 43 L = 1,15
43 EXSCOR(K,L) = 0.0
42 CONTINUE

CPVZ = 0.0

DO 44 N = 1,M
I = IM(N)
J = JM(N)

```

```

RX = RSX(I,J)
RY = RSY(I,J)
CALL DEVIAT (RX,RY,RXY,VZ,DX,DY,SDS,PVZ)
DO 45 K = 1,15
DO 46 L = 1,15
A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5
EXSCOR(K,L) = EXSCOR(K,L) + A
46 CONTINUE
45 CONTINUE
CPVZ = CPVZ + PVZ
44 CONTINUE

SUM = 0.0
SS = 0.0
COV1 = 0.0

RX = COS(ATAN((SIN(ALF1F2)*SIZF2)/(SIZF1 + COS(ALF1F2)
&*SIZF2)))
RY = SIN(ARCOS(RX))
RXYC = 0.0
CALL DEVIAT (RX,RY,RXYC,VZ,DX,DY,SDS,PVZ)
DO 48 K = 1,15
DO 49 L = 1,15
A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5
COV1 = COV1 + A*EXSCOR(K,L)
49 CONTINUE
48 CONTINUE
COV1 = COV1/225.0 - 0.25*XM
DO 50 K = 1,15
DO 51 L = 1,15
SUM = EXSCOR(K,L) + SUM
SS = EXSCOR(K,L)**2 + SS
51 CONTINUE
50 CONTINUE
VEX = SS/225.0 - (SUM/225.0)**2 + 0.01*CPVZ
SDEX = SQRT(VEX)
R1 = COV1/(0.1*SDEX)
R = R1/(XM -(XM - 1.0)*R1)
R1C = (100.0*R)/(1.0 + 99.0*R)

DO 55 M = 1,100
XMAX = -500.0
DO 56 I = 1,15
DO 57 J = 1,15
IF(XMAX - COVSYN(I,J)) 58,57,57
58 XMAX = COVSYN(I,J)
XMOX = XMAX
IM(M) = I
JM(M) = J
57 CONTINUE
56 CONTINUE

```

```

      I = IM(M)
      J = JM(M)
      COVSYN(I,J) = -.500.0
55 CONTINUE

      M = 101
      DO 59 I = 1,15
      DO 60 J = 1,15
      IF(XMOX - COVSYN(I,J)) 61,61,60
61 IM(M) = I
      JM(M) = J
60 CONTINUE
59 CONTINUE
      M = M - 1
      XM = M

      DO 65 K = 1,15
      DO 66 L = 1,15
66 EXSCOR(K,L) = 0.0
65 CONTINUE

      CPVZ = 0.0

      DO 67 N = 1,M
      I = IM(N)
      J = JM(N)
      RX = RSX(I,J)
      RY = RSY(I,J)
      CALL DEVIAT (RX,RY,RXY,VZ,DX,DY,SDS,PVZ)
      DO 68 K = 1,15
      DO 69 L = 1,15
      A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5
      EXSCOR(K,L) = EXSCOR(K,L) + A
69 CONTINUE
68 CONTINUE
      CPVZ = CPVZ + PVZ
67 CONTINUE

      SUM = 0.0
      SS = 0.0
      COV2 = 0.0

      RX = COS(ATAN((SIN(ALF1F2)*SIZF2)/(SIZF1 +
&COS(ALF1F2)*SIZF2)))
      RY = SIN(ARCOS(RX))
      RXYC = 0.0
      CALL DEVIAT (RX,RY,RXYC,VZ,DX,DY,SDS,PVZ)
      DO 70 K = 1,15
      DO 71 L = 1,15
      A = (DX*Z(K) + DY*Z(L))*0.1/SDS + 0.5

```

```

    COV2 = COV2 + A*EXSCOR(K,L)
71 CONTINUE
70 CONTINUE
    COV2 = COV2/225.0 - 0.25*XM
    DO 72 K = 1,15
    DO 73 L = 1,15
    SUM = EXSCOR(K,L) + SUM
    SS = EXSCOR(K,L)**2 + SS
73 CONTINUE
72 CONTINUE
    VEX = SS/225.0 - (SUM/225.0)**2 + 0.01*CPVZ
    SDEX = SQRT(VEX)
    R2 = COV2/(0.1*SDEX)
    R = R2/(XM - (XM - 1.0)*R2)
    R2C = (100.0*R)/(1.0 - 99.0*R)

    WRITE(6,105) RTEST, RITEM, TITMF1, ALF1F2, SIZF1, SIZF2
105 FORMAT(' ', 'RTEST = ', F6.3, 2X, 'RITEM = ', F6.3, 2X,
&'TITMF1 = ', F6.3, 2X, 'ALF1F2 = ', F7.3, 2X, 'SIZF1 = ',
&'F7.3, 2X, 'SIZF2 = ', F7.3)

    WRITE(6,101) R1, R1C, R2, R2C
101 FORMAT(' ', 'R1 = ', F6.3, 2X, 'R1C = ', F6.3, 2X, 'R2 = ',
& F6.3, 2X, 'R2C = ', F6.3)

    GO TO 98
99 CONTINUE
    STOP
    END

```

```

SUBROUTINE DEVIAT (RX,RY,RXY,VZ,DX,DY,SDS,PVZ)

    IF(RX) 1,2,2
1  SRX = -1.0
    GO TO 3
2  SRX = 1.0
3  IF(RY) 4,5,5
4  SRY = -1.0
    GO TO 19
5  SRY = 1.0

19 XK = (RY - RXY*RX)/(RX - RXY*RY)
    XK2 = XK**2
    D = (1.0 + XK*RXY)**2 - RX**2*(1.0 + 2.0*XK*RXY +
&XK2)

    IF(D - 0.0001) 20,20,6
20 IF(D + 0.0001) 6,9,9
9  IF(RX) 10,11,10
11 SDX = 0.0

```



```

SDY = 1.0
VS = 0.0
PVZ = 0.0
GO TO 12

10 IF(RY) 13,15,13
15 SDX = 1.0
SDY = 0.0
VS = 1.0
PVZ = 0.0
GO TO 12

13 VX = 1.0
VY = XK2*VX
SDX = SQRT(VX)
SDY = SQRT(VY)
VS = VX + VY + 2.0*SDX*SDY*RXY
PVZ = 0.0
GO TO 12

6 VX = (RX**2*VZ)/((1.0 + XK*RXY)**2 + RX**2*(1.0 +
&2.0*XK*RXY + XK2))
VY = XK2*VX
SDX = SQRT(VX)
SDY = SQRT(VY)
VX = VX + VY + 2.0*SDX*SDY*RXY
PVZ = VZ/VS

12 SDS = SQRT(VS)
DX = SRX*SDX
DY = SRY*SDY
RETURN
END

```

data

GLOSSARY

Each definition in this glossary is followed by the number of the page in chapter III on which the variable is discussed.

ALF1F2 the angle between the vector representing factor 1 and the vector representing factor 2 (31)

COVEX(I,J) the covariance of an item with the whole test (45)

COVSYN(I,J) the synthetic covariance of an item with the job factors (49)

CPVZ the cumulative sum of PVZ for all items selected for a particular test (51)

DX a scaling factor; multiplied by the z-scores, it generates a random variable X. It is used in the process of generating expected scores for a particular item (38)

DY a scaling factor; the analogue to DX used to generate a random variable Y (38)

EXSCOR(I,J) the expected score of a person whose standing with respect to factors 1 and 2 is indicated by the Ith and Jth z-scores, respectively (35). This Fortran variable is also used to construct a criterion matrix (50).

PVZ the fraction of the variance of scores on a particular item which can be attributed to error; error variance/total variance (51)

RITEM the reliability of a single item for a test of one hundred equally reliable items whose overall reliability is given by RTEST. The value of this variable is determined by RTEST; the inverse cosine of this variable is taken to be a "typical" item-vector angle (32)

RSX(I,J) the correlation of an item with factor 1 (36)

RSY(I,J) the correlation of an item with factor 2 (36)

- RTEST the reliability of a hypothetical test consisting of one hundred items. This value is postulated in order to generate a reasonable "typical" item spread (32)
- RX an equivalent of $RSX(I,J)$ used in the subroutine DEVIAT
- RXY the correlation of the two job factors (36)
- RY an equivalent to $RSY(I,J)$ used in the subroutine DEVIAT
- R1 the validity coefficient of the test selected by an internal criterion (52)
- R1C R1 corrected to represent a test of one hundred items (53)
- R2 the validity coefficient of the test selected by synthetic analysis (52)
- R2C R2 corrected to represent a test of one hundred items (53)
- SDEX the standard deviation of an expected score matrix; the square root of VEX (52)
- SDS the standard deviation of a three-dimensional matrix of values representing scores of a large number of persons on a particular item (52)
- SIZF1 the length of the vector representing factor 1. This value is, in effect the relative importance of factor 1 to factor 2
- SIZF2 the length of the vector representing factor 2
- TITMF1 the overall direction of the item-vectors; the angle between the vector sum of the item-vectors and the vector representing factor 1 (31)
- TYPAL a "typical" angle between an item-vector and the sum of the item-vectors (32)
- VEX the variance of an expected score matrix corrected to include error variance (52)
- VZ the variance of the z-scores (36)

XITMF1(I) the angle between the projection of an item-vector upon the plane of the factor-vectors and the vector representing factor 1 (31)

XITMF2(I) the angle between the projection of an item-vector upon the plane of the factor-vectors and the vector representing factor 2

XK represents the variable A defined by equation (29)
(38)

Z(I) one of fifteen z-scores equivalent to uniformly distributed percentile ranks (28)

ZITMFP(I) the angle between an item-vector and the plane of the factor-vectors (31)

APPROVAL SHEET

The Dissertation submitted by Jerome D. Lehnus has been read and approved by members of the Department of Foundations of Education.

The final copies have been examined by the director of the Dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the Dissertation is now given final approval with reference to content and form.

The Dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date

5/17/76

Signature of Advisor

Jack A. Kavanagh