ACT RESEARCH REPORT

No. 35

35

August, 1970

AN ANALYSIS OF SPATIAL CONFIGURATION AND ITS APPLICATION TO RESEARCH IN HIGHER EDUCATION

Nancy S. Cole
James W. L. Cole

PUBLISHED BY THE RESEARCH AND DEVELOPMENT DIVISION

THE AMERICAN COLLEGE TESTING PROGRAM

P. O. BOX 168, IOWA CITY, IOWA 52240

Abstract

An analysis of the spatial configuration of variables in a multivariate system is presented. The purpose of the analysis is to make clearer the relationships among the variables by locating them in a minimally-dimensioned space. Similarly, individuals are located in the smaller space and related to each other on the basis of the variables measured.

The analysis is then used to locate some colleges on a planar surface on the basis of variables given by Astin. In the configuration of colleges in the plane, a college is described in terms of its relative orientation to several educational aspects and the resulting single point location is suggested as a valuable alternative to profile analysis.

AN ANALYSIS OF SPATIAL CONFIGURATION AND ITS APPLICATION TO RESEARCH IN HIGHER EDUCATION

Nancy S. Cole[1]
James W. L. Cole[2]

Measurement instruments in the social sciences are often multivariate. Rarely, however, are the variables independent. Typically more variables are used than the actual dimensionality of the data suggests. Such practice appears justified when the variables are measures of meaningful characteristics but the dimensions of a smaller dimensioned space are not similarly meaningful or conducive to direct measurement. Even when justified, the greater dimensionality increases the difficulty of clinical and research use of the data. Therefore, the specification of the variables within a minimally-dimensioned space would seem to provide a useful simplification.

The primary motivation for this analysis of spatial configuration is to provide a method for understanding relationships among the variables. The information necessary for such understanding is contained in the correlation matrix, but it is quite difficult to interpret the correlations in the matrix simultaneously. The analysis overcomes this difficulty by often achieving a visual representation of the variables which can be of considerable value in understanding the relationships among the variables.

While using dimension reduction techniques of factor analysis and multidimensional scaling, this analysis is not intended as a method for identifying a smaller number of variables in a system. Thus, the use of factor analysis to replace a large number of variables with a few factors has quite a different motivation from ours. This analysis reduces the dimensionality of the space in which the variables are imbedded but retains the variables. If there were too many variables in the system before the analysis, there will still be too many after the analysis. The purpose is rather to present the variables in a reduced space in which their relationships can be more easily conceived.

A secondary asset of the procedure is that it provides a representation of important aspects of some kinds of profile data with the result that these aspects are more easily and meaningfully evaluated than they are in profile form.

The first portion of the paper presents the mathematical formulation of the method. An illustration of its use and discussion of its advantages and disadvantages in application are presented in the second part of the paper. Readers interested in the application to data may prefer to start with the section beginning on page 5 and use the first section as a reference for the details of the procedure.

---

[1]The authors are indebted to Harold Bechtoldt of the University of Iowa and Mark Applebaum of the University of North Carolina for references and encouragement early in the project. Douglas R. Whitney of the University of Iowa and Gary R. Hanson of ACT made useful contributions at several stages of the development. The critical readings of the manuscript by Robert A. Forsyth of the University of Iowa and E. James Maxey and Leo A. Munday, both of ACT, were very helpful.

[2]James W. L. Cole is Assistant Professor of Statistics at the University of Iowa.

component scores (Harman, 1960, p. 360),

$$\underset{p \times 1}{\underline{z}_j} = \underset{p \times p \times 1}{A\,\underline{f}_j} . \tag{1}$$

Then $\underline{f}_j$ may be found by computing

$$\underline{f}_j = A^{-1}\,\underline{z}_j , \tag{2}$$

when A is of rank p. When the rank of A is essentially $q < p$ (cf. footnote 4), A is taken to be a $p \times q$ matrix and we make the working definition,

$$\underline{f}_j = (A'A)^{-1} A'\underline{z}_j . \tag{3}$$

Since A'A is the diagonal matrix of the characteristic roots, equation (3) gives an easy computational method for either the full or deficient rank case and will be used throughout the remainder of the paper.

The scores $\underline{f}_j$ then locate an individual observation in the p-space of the principal components, the same p-space in which the variables are located. Since both variables and individuals are plotted in the same p-space, it is important that the meaning of this operation be made clear as variables and individuals are rather different things.

One way to understand the results of such a combination is to ask: "What individual score vector $\underline{z}$ would have a corresponding component score vector $\underline{f}$ equal to a given variable point in the p-space?" In other words, recalling that the point in p-space representing the i-th variable is just the i-th row of A or equivalently the i-th column of A', say $\underline{a}_i$, we want to find the vector $\underline{z}^*_i$ such that

$$\underline{z}^*_i = A\,\underline{a}_i . \tag{4}$$

But by the rules of matrix multiplication, $\underline{z}^*_i$ is the i-th column of AA' which is known to equal R. That is to say

$$\underline{z}^*_i = \underline{r}_i , \tag{5}$$

the i-th column of R.

That a variable corresponds to a $\underline{z}$ score equal to a column of R has some intuitive justification. A score on variable j would be expected to lead to a similar score on variable k if j and k are highly correlated. Thus, if the score on variable j is one, the score on variable k equal to r_{jk} is expected simply because of the relation between the variables.

It will be convenient later in the paper if we also find the $\underline{z}$ corresponding to the centroid of the variables in p-space, the mean of the rows of A. For $\underline{a}'_i$, the i-th row of A,

$$\bar{\underline{a}} = (1/p) \sum_{i=1}^{p} \underline{a}_i . \tag{6}$$

Then $\underline{z}^*_m$, the $\underline{z}$ score corresponding to the mean, is given by

$$\begin{aligned} \underline{z}^*_m &= A\,\bar{\underline{a}} \\ &= (1/p) \sum_{i=1}^{p} A\,\underline{a}_i \\ &= (1/p) \sum_{i=1}^{p} \underline{r}_i \\ &= \bar{\underline{r}} . \end{aligned} \tag{7}$$

Thus, the $\underline{z}$ score corresponding to the centroid of the variable vectors is the average of the rows or columns of the correlation matrix.

Step 3: Second Stage Principal Components Analysis

Step 3 begins the second stage of the analysis of spatial configuration. In stage one the variables and individual observations were located in the p-space of the first stage principal components. Now in stage two the problem is to reduce the dimensionality of the p-space to a smaller space in which the relationships among the variables may be more easily understood. Note that we are not discussing the reduction of dimensionality of the original observations as that could have been accomplished, if desired, in the first stage of the analysis. This second stage deals only with the scatter of the variable points about the component axes.

To accomplish a dimension reduction of the variability of the variable points, we first compute the covariance matrix, S, of the p component dimensions over the p variables,

$$S = (1/p)A'A - \bar{\underline{a}}\,\bar{\underline{a}}' . \tag{8}$$

analysis presented here. It has been common in factor analysis to plot variables on the factors computed, often as a step leading to a rotation of factor axes (Thurstone, 1947; Thomson, 1951; Guilford, 1954; and others). In addition, one specific method called the method of extended vectors treats test vectors in the space of the factor axes by extending the vectors to a plane perpendicular to the first factor, or on a unit sphere (Guilford, 1954, p. 514). The location of the variable points on this plane or sphere will often be similar to the configuration of variables achieved in our analysis.

A second factor analysis procedure bears some resemblance to our method though its purpose is entirely different. This is the so-called second order factoring. Second order factoring (Thurstone, 1947; Guilford, 1954) is the procedure of factoring the correlation matrix of nonorthogonal factors with the purpose of discovering more basic dimensions, often in search of a general factor. At the second stage of our analysis we factor the covariance matrix of the components computed only over the p variables, and the second stage principal components analysis is used strictly as a procedure to fit a smaller space to the points, not to discover any factors.

One offshoot of factor analysis and scaling which bears certain resemblances to this method is the work of Guttman on the radex (Guttman, 1954, 1965). Guttman is concerned with relationships among variables—specifically mental tests—in the form of linear or circular relationships. Something like Guttman's simplex or circumplex ordering of variables is often the result of our configural analysis. Guttman's smallest space analysis (Guttman, 1968) is a nonmetric approach definitely in the spirit of our procedure.

An Application of the Configural Analysis to Measures of Colleges

In his book, *Who Goes Where to College?*, Astin (1965) presented student input data for 1,015 four-year colleges and universities. After computing factors for a sample of colleges with extensive data available, Astin used public sources of data to estimate the same factors for the 1,015 institutions. The five estimated student input factors were given the following names and interpretations by Astin (1965, pp. 54-55):

1. Intellectualism (INT). An entering student body with a high score would be expected to be high in academic aptitude (especially mathematical aptitude) and to have a high percentage of its students pursuing careers in science and planning to go on for the Ph.D. degrees.

2. Estheticism (EST). An entering student body with a high score would tend to have a high percentage both of students who achieved in literature and art during high school and of students who aspire to careers in these fields.

Step 3

The covariance matrix, S, of the five components over the five variables is

$$S = \begin{bmatrix} .1599 & & & & \\ -.1481 & .2442 & & & \\ -.0282 & -.0141 & .1396 & & \\ -.0147 & -.0074 & -.0014 & .0433 & \\ -.0305 & -.0153 & -.0029 & -.0028 & .0362 \end{bmatrix}. \quad (13)$$

Table 3 gives the results of a principal components analysis on S. The first two dimensions in the second stage analysis account for 81.2% of the trace. Thus, we know that the deviations of the variable points from their centroid is almost contained in a space of two dimensions (i.e., a plane). Because of the value of providing a visual representation, we may let k = 2.

Table 3

Second Stage Principal Components Analysis

Components:	1	2	3	4
Roots:	0.356	0.149	0.073	0.044
Percent Trace:	57.2	24.0	11.7	7.1

Loadings:

1	−.3621	−.0909	.1433	.0047
2	.4742	−.0826	.1120	.0035
3	.0161	.3655	.0756	.0019
4	.0057	.0128	−.0864	.1888
5	.0117	.0259	−.1640	−.0921

Step 4

For B, the 5 x 2 portion of the loading matrix outlined in Table 3,

$$(B'B)^{-1}B' = \begin{bmatrix} -1.0161 & 1.3306 & 0.0451 & 0.0160 & 0.0329 \\ -0.6077 & -0.5522 & 2.4446 & 0.0857 & 0.1733 \end{bmatrix}. \quad (14)$$

Then the locations of the five variables in two-space can be computed by equation (11). The result, H, is given below.

$$H = \begin{bmatrix} \text{INT} & \text{EST} & \text{STA} & \text{PRA} & \text{MAS} \\ 0.2171 & 1.6475 & 0.2251 & -0.9966 & -1.0931 \\ 0.2447 & 0.6287 & -1.6427 & 1.2652 & -0.4958 \end{bmatrix} \quad (15)$$

The variables can be plotted on the plane to obtain a pictorial representation of the relationships among them. Figure 1 gives that representation.

Step 5

The projection matrix, say P, for locating an individual observation vector (a college) on the plane is computed as in equation (12).

$$P = (B'B)^{-1}B'(A'A)^{-1}A'$$

$$= \begin{bmatrix} 0.1218 & 0.9246 & 0.1264 & -0.5593 & -0.6135 \\ 0.3272 & 0.8409 & -2.1972 & 1.6922 & -0.6631 \end{bmatrix}. \quad (16)$$

Premultiplying by P a college's vector of scores (or a college group mean vector) standardized by $\underline{\Gamma}$ (cf. equation (12)) gives the college's (or group's) location on the plane.

This relatively small importance of the INT variable illustrates one of the peculiar aspects of this form of analysis. The analysis will identify a variable such as INT which does not contribute to the locating of colleges in orientation regions because it is nearly equally correlated with variables which define the regions. The other side of the same coin is the fact that the analysis does *not* reflect differences between colleges on that variable, a point to be remembered in subsequent comparisons between colleges. Thus, on the one hand the analysis reveals the fact that colleges with each of the other four orientations may be either high or low in INT and that no orientation is systematically higher or lower than others. That is to say, INT, with the strong influence of academic aptitude and educational plans, cuts across other orientation variables. On the other hand, the magnitude of the INT variable, which may be of interest in itself, is largely ignored in this analysis.

Colleges, some examples

Given the limitation of the analysis for this particular data, let us consider what the analysis does tell us about colleges' orientations. We will do this by comparing the discussion of college profiles and conclusions of Astin with our analysis and conclusions.

Consider first two profiles which Astin (1965, p. 86) used to illustrate his data, those of Rice University and Princeton University. An adaptation of the profiles is reproduced in Figure 2 for the five variables under consideration. Astin discussed three kinds of interpretation that can be made: (1) comparison of scores within an institutional profile; (2) comparison with institutions in general; and (3) comparison of specific institutions with each other. Each type of interpretation has a parallel based on the configural analysis, but as will be seen there are definite differences between the types of conclusions that can be drawn from the two types of analysis.

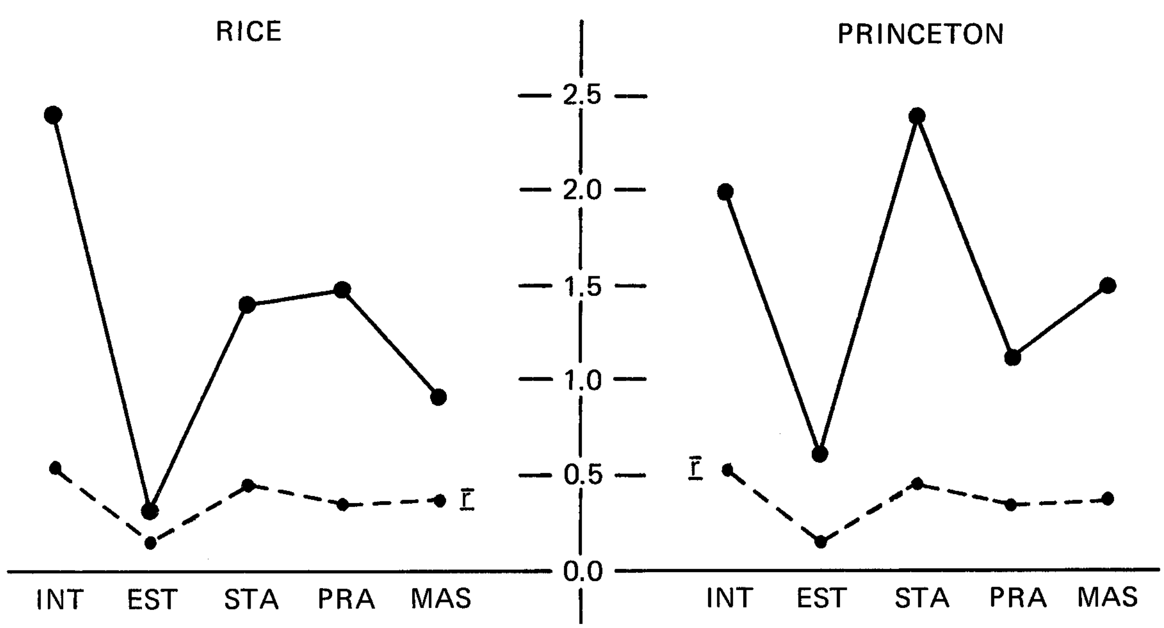

Figure 2. Profile of two colleges on five Astin variables.

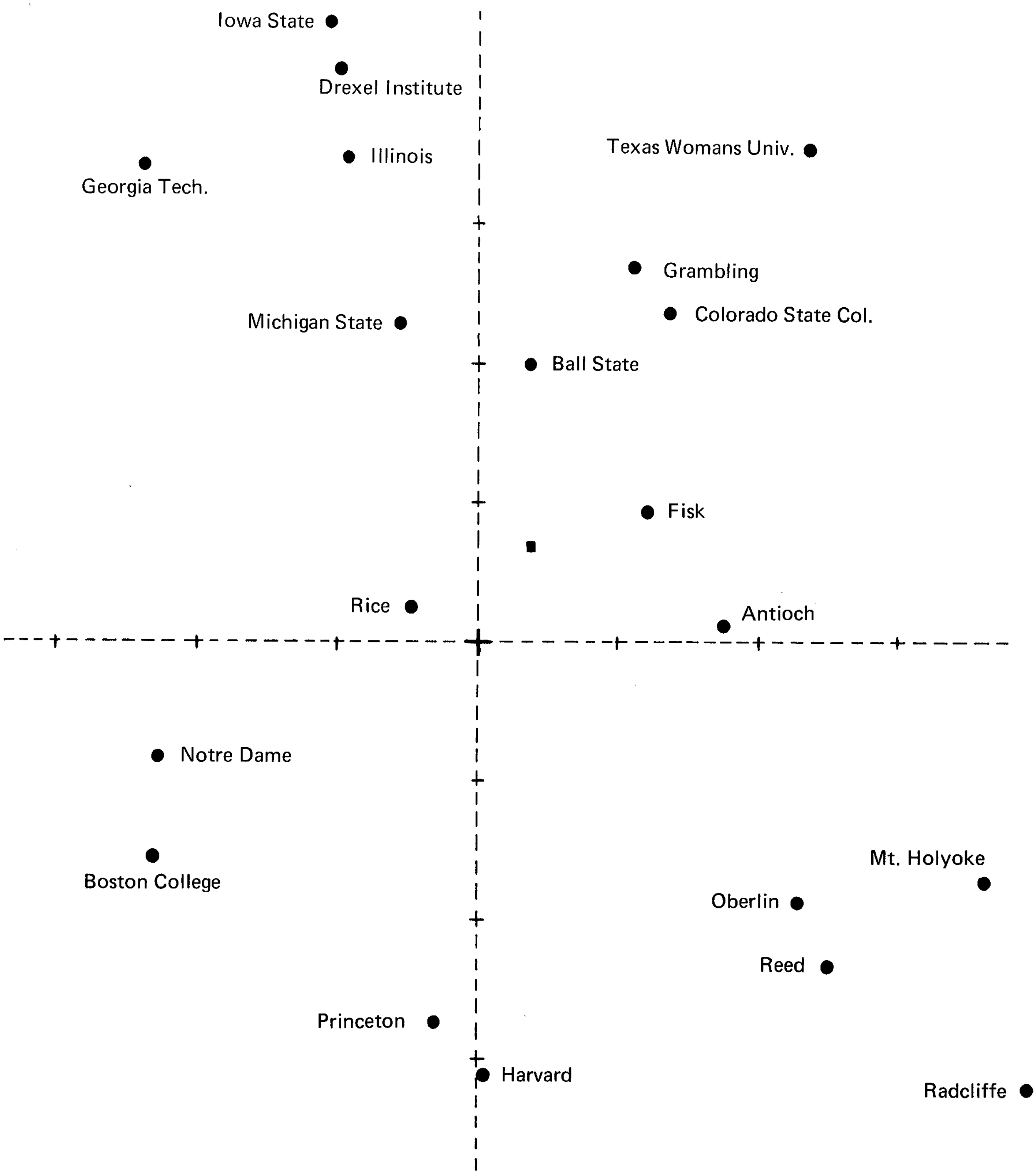

Figure 3. Location of colleges on the plane.

ing, etc.) rather than Social fields. As public institutions they lack a strong Status domination, have some Masculinity pull, and as a result are located generally to the upper left. Teacher colleges and some former teacher colleges are located to the upper right, reflecting the relative strength of their Social and Artistic orientations. Note that two predominantly Negro colleges fall in this quadrant also, suggesting a similarity in orientation with the teacher colleges. Several small, elite private women's colleges and artistically-oriented liberal arts colleges are located in the lower right quadrant. These colleges show STA and EST orientations. Elite private universities with more MAS pull and less Social and EST pull are to the lower left.

Types of Colleges

The observations made above on the basis of the sample of institutions in Figure 3 can be made more systematically by computing means for groups of colleges of the same type. Fifteen types of colleges were considered and these types are listed in Figure 4 along with the number of colleges included in each group.

The locations of college type means in Figure 4 support the observations made about where types of colleges fall. It is interesting to note the strong STA orientation (high SES and Enterprising emphasis) in the most selective and smallest colleges. Of course, this result comes as no surprise. It is mainly the public institutions which offset the Status domination in favor of a heavy Pragmatic orientation with emphasis on Realistic vocational choices in the universities and Social and Artistic emphasis in the teacher colleges. Thus, clearly the public institutions in this country have more egalitarian orientations than the private colleges as well as technical and social orientations rather than business orientations.

One final point to be made is that there seem to be consistent and meaningful distinctions which can be made about colleges on the basis of this analysis of spatial configuration. Recall that these distinctions have made very little use of the differences in academic ability of the students. Thus, it seems that one can discuss the diversity of orientations in American colleges and universities in this framework with no reference to or implication of corresponding differences in academic ability of the students.

Other Applications

One important group of instruments for which the analysis of spatial configuration seems especially suited is the interest inventories. Here a vector of scores is interpreted in terms of the relative orientation of the individual to different interest patterns. Often measures of absolute degree of interest in any area covered by the instrument are complicated by such things as response sets and are therefore not of greatest importance.

A configural analysis of Holland's Vocational Preference Inventory (Cole, Whitney, and Holland, in press) has been performed. The analysis was helpful in relating the VPI variables to each other, in "typefying" an individual by his location on the plane, and in locating occupational groups on the plane.

Because of the way the configural analysis concerns the dimensions on which variables differ, it seems especially appropriate as a way to study the relationships of scales in an instrument. In a paper concerned with differential validity in a battery of tests, it was informative to consider the pattern of the test variables on the second and third principal components (Cole, 1969). A similar configuration would be obtained by this analysis. When variables are too close together they may be measuring too similar a concept. Areas may appear in which a measure is called for but has not been included.

One interesting possibility is to use the analysis at the level of scale construction in order to discover the dimensions on which the items on a scale differ.[6] Then one can judge which dimensions of item differences seem to be important ones and which are to be eliminated. When two scales are closely related, items from both may be analyzed simultaneously in order to discover overlapping items responsible for the similarity.

[6] Gary R. Hanson suggested this use of the configural analysis for item analysis. Preliminary analyses by Hanson suggest that the analysis can be very useful in this way.

References

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: John Wiley, 1958.

Astin, A. W. *Who goes where to college?* Chicago: Science Research Associates, 1965.

Cole, N. S. *Differential validity in the ACT tests.* Research Report No. 30. Iowa City: The American College Testing Program, 1969.

Cole, N. S., Whitney, D. R., & Holland, J. L. A spatial configuration of occupations. *Journal of Vocational Behavior,* in press.

Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Illinois: The Free Press, 1954, 258-348.

Guttman, L. The structure of interrelations among intelligence tests. *Proceedings of the 1964 invitational conference on testing problems.* Princeton: Educational Testing Service, 1965, 25-36.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika,* 1968, *33*, 4, 469-506.

Harman, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology,* 1933, *24,* 417-441, 498-520.

Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine,* 1901, *2* (sixth series), 559-572.

Thomson, G. H. *The factorial analysis of human ability.* Boston: Houghton Mifflin, 1951.

Thurstone, L. L. *Multiple factor analysis.* Chicago: University of Chicago Press, 1947.

Torgerson, W. S. *Theory and methods of scaling.* New York: John Wiley, 1958.

ACT Research Reports

This report is the thirty-fifth in a series published by the Research and Development Division of The American College Testing Program. The first 26 research reports have been deposited with the American Documentation Institute, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Photocopies and 35 mm. microfilms are available at cost from ADI; order by ADI Document number. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress. Beginning with Research Report No. 27, the reports have been deposited with the National Auxiliary Publications Service of the American Society for Information Science (NAPS), c/o CCM Information Sciences, Inc., 22 West 34th Street, New York, New York 10001. Photocopies and 35 mm. microfilms are available at cost from NAPS. Order by NAPS Document number. Advance payment is required. Printed copies may be obtained, if available, from the Research and Development Division, The American College Testing Program. The reports are indexed by the *Current Contents, Education* Institute for Scientific Information, 325 Chestnut Street, Philadelphia, Pennsylvania 19106.

The reports since January 1968 in this series are listed below. A listing of previous reports is included in each of several items published by The American College Testing Program: *Your College Freshmen* (pp. 158-160), *Your College-Bound Students* (pp. 107-109). A complete list of the reports can be obtained by writing to the Research and Development Division, The American College Testing Program, P. O. Box 168, Iowa City, Iowa 52240.

No. 23 *The Effects of Selecting College Students by Various Kinds of High School Achievement,* by L. L. Baird & J. M. Richards, Jr. (ADI Doc. 9955; photo, $3.75; microfilm, $2.00)

No. 24 *Do They Do What They Say They Will Do?* by S. W. Lutz (ADI Doc. 9988; photo, $5.00; microfilm, $2.25)

No. 25 *Changes in the Vocational Plans of College Students: Orderly or Random?* by J. L. Holland & D. R. Whitney (ADI Doc. 10051; photo, $3.75; microfilm, $2.00)

No. 26 *The Flow of High School Students to Schools, Colleges, and Jobs,* by L. L. Baird & J. L. Holland (ADI Doc. 10053; photo, $6.25; microfilm, $2.50)

No. 27 *Forecasting Academic Success in Specific Colleges,* by D. P. Hoyt (NAPS No. 00254; photo, $3.00; microfilm, $1.00)

No. 28 *A Description of Graduates of Two-Year Colleges,* by L. L. Baird, J. M. Richards, Jr. & L. R. Shevel (NAPS No. 11306; photo, $3.00; microfilm, $1.00)

No. 29 *An Empirical Occupational Classification Derived from a Theory of Personality and Intended for Practice and Research,* by J. L. Holland, D. R. Whitney, N. S. Cole & J. M. Richards, Jr. (NAPS No. 00505; photo, $3.00; microfilm, $1.00)

No. 30 *Differential Validity in the ACT Tests,* by N. S. Cole (NAPS No. 00722; photo, $3.00; microfilm, $1.00)

No. 31 *Who Is Talented? An Analysis of Achievement,* by C. F. Elton & L. R. Shevel (NAPS No. 00723; photo, $3.00; microfilm, $1.00)

No. 32 *Patterns of Educational Aspiration,* by L. L. Baird (NAPS No. 00920; photo, $3.00; microfilm, $1.00)

No.33 *Can Financial Need Analysis Be Simplified?* by M. D. Orwig & P. K. Jones (NAPS No. not available at this time.)

No. 34 *Research Strategies in Studying College Impact,* by K. A. Feldman (NAPS No. not available at this time.)