DOCUMENT RESUME

ED 193 301 TM 800 616

AUTHOR	Thompson, Bruce; Frankiewicz, Ronald G.
TITLE	Estimating Reliability of Factor Analytic Results.
PUB DATE	80
NOTE	16p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 7-11, 1980).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Attitude Measures; Error of Measurement; *Factor Analysis; *Measures (Individuals); *Test Reliability; Weighted Scores
IDENTIFIERS	*Alpha Coefficient; Dressel (P L); *Estimation (Mathematics)

ABSTRACT

A procedure for estimating reliability in a factor analytic context, when reliability of the extracted factors is not an emphasis, is identified. The procedure is an extension of Dressel's work and might be applied in attitude measurement. It assesses how homogeneous the weighted original item responses are, when they are scored for pattern congruity. Four data sets generated to simulate different sample sizes, numbers of variables, numbers of variables per factor, and proportions of randomly generated error, were analyzed with alpha factor analysis followed by rotation to the varimax criterion. Subsequently, both types of consistency estimates were calculated. Results indicate that the two indices are sensitive to different data characteristics. The alpha coefficients are primarily sensitive to the variance each factor explains, as reflected by the number and magnitude of the salient pattern coefficients associated with each factor. The Dressel coefficients, however, are sensitive to the proportion of error embedded in the original data, and are less sensitive to the number of salient items associated with each factor. (RL)

Estimating Reliability of Factor Analytic Results

Bruce Thompson

University of New Orleans

Ronald G. Frankiewicz

University of Houston

Paper presented at the annual meeting of the National Council on Measurement in Education, 1980. The authors wish to indicate that they contributed equally to the development of this paper.

Educational researchers have long recognized the usefullness of factor analytic procedures. Before widespread access to computers was achieved, factor analysis was infrequently performed, because factor analytic techniques require rather forbidding calculations. The promulgation of computers brought a corresponding increase in the use of factor analysis, however. In fact, for a brief period researchers almost seemed to factor routinely virtually any data which they collected. This abuse abated somewhat with an increasing recognition that knowledge of the dimensions underlying variables is not per se valuable, and several didactic demonstrations that identified factors are not necessarily "true" primary dimensions. For example, Overall (1964) factor analyzed book measurements and isolated the dimensions of size, obesity, and squareness, rather than the more commonly accepted primary dimensions of heighth, width, and thickness.

Today factor analytic methods are more sensibly applied only in the context of specific theory regarding the dimensionality of measures (cf. Guilford, 1967), as a valuable tool for assessing validity (Nunnally, 1967, pp. 100-101, Thompson, in press), and in some experimental research as a prelude to various ANOVA procedures and their analogues (Morrow and Frankiewicz, 1979). Factor analysis can be practically employed to examine the dimensions underlying several of the "factorable entities" discussed by

Cattell (1966). Thus researchers have employed factor analysis to factor variables (Thompson and Rucker, in press), to factor people (Korb and Frankiewicz, 1976), and to factor occasions of measurement (Jones, Miller, and Thompson, in press).

However, as Cronbach (1951, p. 297) has emphasized, "no factor analysis can be interpreted without some appropriate estimate of the magnitude of the error of measurement." Researchers typically rely on alpha factor analysis (Kaiser and Caffrey, 1965) to generate these estimates. Alpha factor analysis assumes that variables rather than persons have have been sampled, and that the researcher is interested in estimating how closely obtained results correspond to the results that would have been obtained if the entire population of variables measuring a construct had been used in place of only a sample of the variables from this universe. Of course, researchers rarely employ every available measure from any universe of measures, because the number of variables which measure most constructs is usually very, very large indeed. Thus, this correspondence estimate is frequently important, because

> there is no practical testing problem where the items in.the test and only the items constitute the trait under examination. We may be unable to compose more items because of out limited skill as testmakers but any group of items in a test of intelligence or knowledge or emotionality is regarded as a sample of items. (Cronbach, 1951, p. 307)

This paper, however, rests upon a different view of educational measurement. Researchers do not always wish to generalize from a sample of measures to the entire universe of possible measures. This occurs, for example, when a given set of measures has been frequently used in past research, and the researcher is primarily interested in generalizing across studies rather than to an abstract population of all possible measures. Clearly, alpha factor analysis is helpful when measurement and theoretical issues are addressed, but the reliability coefficients which it generates are couched in a perspective which may not always be appropriate.

The inquiry reported here was conducted in order to identify a procedure for estimating reliability in a factor analytic context when reliability of the extracted factors is not an emphasis. The inquiry was also conducted for the further purpose of estimating how adequately the procedure performs. The procedure is an extension of Dressel's (1940) work, and might be applied in attitude measurement.

Psychometrics of the Methods

"Alpha" is the name which Cronbach (1951) gave to a coefficient which is sensitive to the internal consistency of data generated by a set of variables. The primary advantage of alpha over split-half consistency estimates is that alpha is unique for any data set while split-half consistency estimates vary over different arbitrary splits of the data. One interpretation of alpha is that it represents the mean of all possible split-half coefficients and is the general case for which the Kuder-Richardson estimate is a special case. More complete discussion of alpha is available elsewhere (Cronbach, 1951, McKennell, 1977, Tyron, 1957).

In the context of factor analysis, Kaiser and Caffrey (1965, p. 6) suggest that alpha is expressed by (1), as a function of the v number of variables in a measure and the eigenvalue (λ) for the j-th factor. Although there is some doubt (McDonald, 1970, pp. 16-20) that this value can always be interpreted as a generalizability coefficient (Cronbach, Gleser, Nanda, and Rajaratnam, 1972), the value is certainly an estimate of factor reliability. Glass (1966) has argued that the results obtained with alpha factor analysis are valuable in any research utilizing fallible measures (always!?). Conceptually, alpha is a weighted average of the intercorrelations of the factored measure's items.

<u>Insert Equation 1 about here</u>

However, an alternative estimate of consistency may be examined in a factor analytic context. This approach is based on an extension of (2) (Dressel, 1940, p. 309), in which a_i is some weight applied to the scores of each n persons on each i items, and p_i is the proportion of the n

persons who "succeeded" on the item. This formula is helpful to researchers who wish to estimate reliability for cognitive measures, which have right-wrong answers, when different questions "count" different amounts.

Insert Equation 2 about here.

In a factor analytic context, $\underline{a}_i$ might be defined as the magnitude of a factor pattern coefficient of item $\underline{i}$ on the $\underline{j}$-th factor. How then should "success" be defined if the measure is attitudinal? Success might be defined by examing the data provided by individual subjects. An intra-subject mean (or median) can be calculated for each person by pooling responses across items. The item $\underline{i}$ response of the $\underline{n}$-th individual might be considered successful if the item varies about the $\underline{n}$-th subject's mean in a manner which reflects the pattern of the salient pattern coefficients for the $\underline{j}$-th factor. Test statistics can be applied to determine pattern coefficient salience, but the identified magnitudes would largely be an artifact of sample size. A more reasonable strategy might involve a "meaningfullness" criterion. This value would vary from one study to the next, but generally an item might be considered salient to a factor if the item had a structure coefficient greater in magnitude than .3. A structure coefficient of .3 indicates that the item shared at least nine percent of its variance with the factor.

Conceptually the proposed extension of Dressel's formula assesses how homogeneous are the weighted original item responses, when they are scored for pattern congruity. It is important to emphasize that the index assesses the weighted consistency of the original responses, and not the reliabilities of factors.

Method & Results

The inquiry involved generation of Monte Carlo-type data for situations involving the Table 1 parameter combinations. Four data sets were generated to simulate different sample sizes, numbers of variables, numbers of variables per factor, and proportions of randomly generated error. All four data sets were analyzed with alpha factor analysis followed by rotation to the varimax criterion. Subsequently, both types of consistency estimates were calculated; these are presented in Table 2.

Insert Tables 1 & 2 about here.

Discussion

Table 2 suggests that, as expected, the two indices are sensitive to different data characteristics. The alpha coefficients are primarily sensitive to the variance each factor explains, as reflected by the number and magnitude of the salient pattern coefficients associated with each factor. This could be problematic, if researchers did not carefully scrutinize the original data for reliability.

This study and several others (cf. Horn and Knapp, 1973) suggest that high alpha coefficients can be produced even by unreliable or random data. The Dressel coefficients, however, are sensitive to the proportion of error embedded in the original data, and are less sensitive to the number of salient items associated with each factor. Consequently, these results suggest that the Dressel type procedure has some promise as an estimate of the weighted pattern congruity of the original data.

Of course, much refinement and further examination must be completed before the procedure can be considered usable. However, if the procedure ultimately proves to be viable, several advantages might be realized. First, the technique could be employed to identify subjects whose responses deviate dramatically from model expectations. Some research (McCollum and Thompson, 1980) suggests that response style and bias errors in attitudinal measures may be much more serious than is generally recognized, and one strategy for minimizing and detecting these influences when ipsative measures can be employed has been identified. But the Dressel type procedure might usefully be employed when measurement is normative. The procedure might be utilized to identify "misfitting" persons in a manner analogous to latent trait theory identification of "misfitting" persons (Wright and Stone, 1979).

Secondly, the technique offers some hope that the consistency of data can be examined in a factor analytic context when the researcher does not wish to generalize to a universe of all possible construct measures. This will be helpful when methods other than alpha factor analysis are utilized, or when the factored matrix of associations is a variance-covariance rather than a correlation matrix (Thompson and Stapleton, in press). Ultimately, the procedure may provide a helpful adjunct to the use of factor reliability (Kaiser and Caffrey, 1965) and invariance (Gorsuch, 1974, p. 309) coefficients.

References

Cattell, R.B. The data box: Its ordering of total resources in terms of possible relational systems. In R.B. Cattell (Ed.), Handbook of multivariate psychology. Chicago: Rand-McNally, 1966.

Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley, 1972.

Dressel, P.L. Some remarks on the Kuder-Richardson reliability coefficient. Psychometrika, 1940, 5, 305-310.

Glass, G.V Alpha factor analysis of infallible variables. Psychometrika, 1966, 31, 545-561.

Gorsuch, R.L. Factor analysis. Philadelphia: W.B. Saunders, 1974.

Guilford, J.P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Horn, J.L., & Knapp, J.R. On the subjective character of the empirical base of the structure-of-intellect model. Psychological Bulletin, 1973, 80, 33-43.

Jones, H.L., Thompson, B., & Miller, A.H. How teachers perceive similarities and differences among various teaching models. Journal of Research in Science Teaching, in press.

Kaiser, H.F., & Caffrey, J. Alpha factor analysis. *Psychometrika*, 1965, *30*, 1-14.

Korb, R., & Frankiewicz, R.G. Strategy for *a priori* selection of judges in a product-centered approach to assessment of creativity. *Perceptual and Motor Skills*, 1976, *42*, 107-115.

McCollum, J., & Thompson, B. *Response error as an external validity threat in evaluation studies*. Paper presented at the annual meeting of the American Educational Research Association, 1980.

McDonald, R.P. The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *The British Journal of Mathematical and Statistical Psychology*, 1970, *23*, 1-21.

McKennell, A.C. Attitude scale construction. In C.A. O'Muircheartaigh & C. Payne (Eds.), *Exploring data structures* (Vol. 1). New York: John Wiley, 1977.

Morrow, J.P., & Frankiewicz, R.G. Strategies for the analysis of repeated and multiple measures designs. *Research Quarterly*, 1979 *50*, 297-304.

Nunnally, J.C. *Psychometric theory*. New York: McGraw-Hill, 1967.

Overall, J.E. Note on the scientific status of factors. *Psychological Bulletin*, 1964, *61*, 270-276.

Thompson, B. Validity of an evaluator typology. *Educational Evaluation and Policy Analysis*, in press.

Thompson, B., & Rucker, R. Students' goals and program preferences. Journal of College Student Personnel, in press.

Thompson, B., & Stapleton, J. A strategy for validating semantic differential referents. Journal of Experimental Education, in press.

Tyron, R.c. Reliability and behaviour domain validity: reformulation and histrocial critique. Psychological Bulletin, 1957, 54, 9-17.

Wright, B.D., & Stone, M.H. Best test design. Chicago: MESA Press, 1979.

TABLE I

Parameter Combinations

N	V	V/Factor[a]	Error Proportion
400	15	5,4,3,2,1	moderate
100	20	5,4,3,2,1	moderate
100	20	5,4,3,2,1	large
60	25	5,4,3,2,1	moderate

[a]For first 15 items before error introduced.

TABLE II

Homogeneity Coefficients

Situation	Factor	N of Salients	Alpha	Dressel
1	I	5	.86	.70
	II	3	.78	.71
	III	2	.72	.47
	IV	4	.54	.59
	V	0	.12	--
2	I	5	.84	.67
	II	3	.74	.70
	III	4	.71	.52
	IV	2	.62	.46
	V	3	.58	.08
	VI	1	.35	.01
	VII	1	.28	.01
	VIII	2	.11	.10
3	I	6	.82	.32
	II	3	.73	.63
	III	2	.69	-.01
	IV	1	.63	.01
	V	2	.61	.37
	VI	1	.51	.01
	VII	2	.30	.11
	VIII	2	.14	-.11
4	I	4	.84	.65
	II	6	.78	.62
	III	3	.74	.46
	IV	4	.68	.34
	V	5	.62	.33
	VI	4	.59	.16
	VII	3	.50	.30
	VIII	1	.34	.02
	IX	2	.21	.12

$$\alpha_j = \left(\frac{v}{v-1}\right)\left(1 - \frac{1}{\lambda_j}\right) \qquad (1)$$

$$r_t = \left(\frac{v}{v-1}\right)\left\{1 - \frac{\Sigma a_i^2 p_i (1-p_i)}{\sigma_t^2}\right\} \qquad (2)$$