DOCUMENT RESUME

ED 205 553 TM 810 432

AUTHOR Thompson, Bruce
TITLE Factor Stability as a Function of Item Variance.
PUB DATE 81
NOTE 14p.: Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attitude Measures: *Factor Analysis: Factor Structure: Psychometrics: *Rating Scales: *Research Problems: Scoring: Test Reliability: *Test Validity
IDENTIFIERS Data Sets: *Item Variance

ABSTRACT

Factor analysis can be a valuable tool for assessing validity: however, the reliabilities of identified factors must be attended to. This study examined the impacts of item variance on factor stability and structure. The attempt was to determine empirically whether unnumbered graphic scales scored using a larger number of scale steps produce more stable factors than do unnumbered graphic scales scored using a smaller number of scale steps. Inservice teachers read and rated their preference of summaries of 16 models of teaching. The scales were rescored using five different numbers of scale steps. Results suggest that researchers need not be concerned that an overly-specific scoring metric will distort the meaning of the identified factors. In addition, it is concluded that unnumbered graphic scales are a promising method of collecting attitudinal data, particularly if the scales are scored with a reasonably large number of scale steps. (Author/GK)

Factor Stability as a Function of Item Variance

Bruce Thompson

University of New Orleans 70122

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, 1981.

Abstract

Factor analysis can be a valuable tool for assessing validity. However, researchers using the method must attend to the reliabilities of identified factors. The study examined the impacts of item variance on factor stability and structure. The results suggest that one way factor reliability can be maximized is by collecting attitudinal data using an unnumbered graphic scale scored with a reasonably large number of scale steps.

The value of factor analysis as an aide to measurement has been demonstrated in numerous studies (Thompson & Miller, in press), although it is also clear that factor analytic techniques can be applied inappropriately (Overall, 1964). Factor analytic methods can be appropriately applied in the context of theory regarding the structure of measures (Guilford, 1967), as a valuable tool for assessing measurement validity (Thompson & Pitts, in press), and in some experimental research as a prelude to various analysis of variance procedures and their analogues (Morrow & Frankiewicz, 1979). 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." Several researchers have attempted to formulate ways to maximize factor reliability because the magnitude of the contribution of factor analytic results is largely a function of factor reliability.

Two general classes of methods for maximizing these reliabilities have been identified. First, efforts can be made to maximize factor reliability <u>after</u> data have been collected, i.e., during the analytic process. Alpha factor analysis (Kaiser & Caffrey, 1965), for example, represents an effort to extract factors which are maximally reliable. Similarly, Hendrickson and Green (1972) have explored the effects which Guttman-weighting of item responses has upon reliability and factor structure. However, it also also possible to maximize factor reliability by attending to reliability considerations <u>during</u> the data collection process. These efforts are generally directed toward maximizing the stability of the entries in the matrix of associations which is to be factored. Although this matrix can take the form of a variance-covariance matrix (Thompson & Stapleton, 1980), in practice most researchers analyze instead an inter-item correlation matrix.

Psychometric Considerations

The stability of indices of association can be maximized in the same way that reliability is most easily maximized, i.e., by attempting to increase the standard deviations of the data. As Gronlund (1976, p. 118) explains, since reliability is maximized "when individuals [or items] tend to stay in the same relative position in a group, from one testing to another, it naturally follows that anything which reduces the possibility of shifting positions in the group also contributes to larger reliability coefficients [or more stable association estimates]. In this case greater differences between the scores of individuals reduce the possibility of shifting positions." At least three strategies for maximizing the stability of indices of association can be identified: include more items on the measure, include more and different kinds of people in the sample, or allow a larger number of responses for each item. Of course, the three strategies each "allow" subjects to generate data which are more spread out, but do not insure that the subjects will actually do so.

Unfortunately, the first two strategies for maximizing the stability of indices of association are not always practical or desirable. For example, longer instruments can contribute to lower response rates or greater experimental mortality. Lengthening an instrument can also lower reliability if fatigue begins to affect results. Similarly, including more or different kinds of people in a sample may be impractical. However, in some cases the number of available response alternatives can be increased without placing an overwhelming burden on subjects.

Response alternatives for attitudinal measures can be presented in at least two ways. Numerical scales provide subjects with a response key, e.g., "SA" = STRONGLY AGREE, "A" = AGREE, "D" = DISAGREE, "SD" = STRONGLY DISAGREE. The subjects then circle or write the appropriate response for each item based upon the identified key. However, numbered graphic scales can also be employed to collect attitudinal data. Graphic scales are frequently employed with semantic differential items, for example, when subjects are asked to mark through a numbered continuum drawn between two bipolar adjectives. Nunnally (1967, p. 520) has suggested that graphic scales should be preferred over numerical scales, because "the presence of a graphic scale probably helps to convey the idea of a rating continuum. Second, the graphic scale should lessen clerical errors in making ratings."

However, a third method can be employed to collect attitudinal data in a non-ipsative manner. Each subject can be asked to draw a line through an unnumbered continuum drawn between two bipolar adjectives. Each subject is instructed to draw the line through the continuum at the point which best represents the desired response. This last measurement alternative might be termed an unnumbered graphic scale and has been employed successfully in previous research (Jones, Thompson & Miller, 1980; Thompson, Borgers & Ward, 1978).

Number of Scale Steps

Each of the three rating scale techniques requires that a choice be made regarding the number of scale steps which will be employed. When numerical or numbered graphic scales are employed the decision must be made before the data are collected, because the scales communicate to the subjects the number of steps which are being utilized. However, when unnumbered graphic scales are employed the researcher has some lattitude in making the choice of the number of scale steps and conceivably could do so even after the data have been collected. In any case, a choice of the number of scale steps must be made at some point for all three methods.

When maximizing the stability of a matrix of associations is a consideration, psychometric theory and previous research (Allison, 1972; Guilford, 1954, pp. 289-291) suggest that the advantage lies with using more rather than fewer scale steps. More steps allow larger standard deviations to be generated so that stability of the association indices is maximized. As Nunnally (1967, p. 521) explains, "it is true that, as the number of scale points increases, the error variance increases, but at the same time, the true-score variance increases at an even more rapid rate." Thus Guilford (1954, P. 291) suggests that "it may pay in some favorable situations to use up to 25 scale divisions." Use of a large number of scale steps only becomes undesirable when subjects become confused or irritated at being confronted with a cognitively overwhelming number of response alternatives.

One advantage of unnumbered graphic scales is that they can be scored using a relatively large number of scale steps. This should result in larger item standard deviations, higher item reliabilities, and finally more reliable factors. The approach is not frustrating to subjects because the number of intervals is not explicitly communicated.

Of course, the approach presumes that the subjects are all making similar and fairly refined judgments. There is currently at least incidental evidence that this presumption is tenable. Various studies (Brown, 1976, 1977, 1978, 1979; Miller, Thompson & Frankiewicz, 1975; Thompson, 1980; Thompson & Miller, 1978) have employed an instrument with unnumbered graphic scales which were subsequently scored using a 15-unit equal-interval scale. The factor structures identified in these studies have been remarkably invariant across both cultural groups and referents. The structures should have been less invariant if the solutions capitalized on measurement error introduced by idiosyncratic scale interpretation or an untenable presumption that refined judgments were being made.

Thus, it is conceivable that larger item standard deviations may be achieved by employing unnumbered graphic scales to collect data and subsequently scoring the scales using a reasonably large number of scale steps. The purpose of this research was to determine empirically whether unnumbered graphic scales scored using a larger number of scale steps produce more stable factors than do unnumbered graphic scales scored using a smaller number of scale steps. The study also examined what impacts scoring metrics have upon comparability of factor structures.

Method

Data were obtained by conducting a secondary analysis of results previously reported by Thompson (1980). The subjects in the previous study were 235 inservice teachers. The subjects read summaries of 16 models of teaching. The validity of the summaries had been previously examined (Jones, Thompson & Miller, 1980). Each subject rated preference for each model of teaching on an unnumbered graphic scale.

For the purposes of this methodological inquiry the scales were subsequently rescored using five different numbers of scale steps. The scales were scored using equal-interval scales with respectively 2, 5, 9, 17, and 33 steps. If factor stability and structure are insensitive to item variance, since the five data sets were generated by the same subjects on the same instrument, the factors produced by the data sets should both be equally reliable and have similar structures.

Results

The data sets were analyzed using alpha factor analysis (Kaiser & Caffrey, 1965). The procedure generates maximally reliable factors. The procedure also produces estimates of the internal consistency reliability of each factor. Although there is some argument regarding the matter (McDonald, 1970, pp. 16-20), it has been suggested that the estimates can also be interpreted as generalizability coefficients. The average standard deviations for the 16 items in the five data sets were respectively .47, 2.94, 3.25, 4.45, and 8.54. The numbers of extracted factors were respectively six, seven, seven, four, and four. All factors with eigenvalues greater than one were extracted.

The alpha coefficients for the factors identified in the analysis are reported in Table 1. These coefficients are estimates of the internal consistency reliability of each factor. Table 1 also reports the cosines of the angles between each factor and the related factor in the solution involving either more or fewer scale steps. These cosines are essentially validity coefficients (Kaiser, Hunka & Bianchini, 1969); they represent the correlation between two factors once solutions have been rotated to a position of best fit. Further explanation and a heuristic example of the procedure have been provided by Gorsuch (1974, pp. 252-257).

<u>Insert Table 1 about here.</u>

Discussion

The cosines presented in Table 1 suggest that the selection of the number of scale steps to be employed in a study can affect the structure of the identified factors. The solutions involving 17 and 33 scale steps were virtually identical. The solutions involving 5 and 9 scale steps were reasonably comparable. Other combinations of scale steps produced factor structures which were dramatically different. The alpha coefficients presented in Table 1 suggest that the factors produced from data involving more scale steps tend to be minimally more reliable.

The fundamental difference in the solutions is that fewer factors were extracted for the data sets involving a larger number of scale steps. Furthermore, the factors with the lowest alphas in each solution were more reliable for solutions involving more scale steps. There are several reasons to believe that the solutions involving more scale steps are more likely to be replicable. The solutions produce fewer factors, and parsimony in number of factors minimizes effects of error and sampling specificity (Peterson, 1965). The solutions for data involving more scale steps also produced factors with more salient items per factor. As Gorsuch (1974, p. 295) explains, "as the number of salient variables per factor increases, the rotational position will be more uniquely determined and will involve less capitalization on chance. The less capitalization on chance [which] occurs, the more replicable the results should be."

It is also important that the solutions for data with 17 and 33 scale steps were so comparable. This result suggests that the researcher need not be extraordinarily worried that an overly-specific scoring metric will distort the meaning of the identified factors. In short, the results suggest that unnumbered graphic scales are a promising method of collecting attitudinal data, particularly if the scales are scored with a reasonably large number of scale steps.

References

Allison, H. K. Rating scale format as it affects ratings in student teaching. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. (ERIC Document Reproduction Service No. ED 152 857)

Brown, R. Teaching style preference as a function of learner intellective and interpersonal dispositions (Doctoral dissertation, University of Houston, 1976). Dissertation Abstracts International, 1976, 37, 4976-A (University Microfilms No. 77-1, 499).

Brown, R. The relationships among student evaluation of teaching, student achievement and student perception of teacher effectiveness. Fresno: California State University, 1977. (ERIC Document Reproduction Service No. ED 133 314)

Brown, R. Effects of perceived teacher-learner style congruency on teacher evaluation. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

Brown, R. Perceptions of teaching/learning style: The mediating process in student evaluation of instruction. CEDR Quarterly, 1979, 12, 16-18.

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

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

Gronlund, N. E. Measurement and evaluation in teaching (3rd ed.). New York: MacMillan, 1976.

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

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

Hendrickson, G. F., & Green, B. F., Jr. Comparison of the factor structure of Guttman-weighted vs. right-only-weighted tests. Paper presented at the annual meeting of the American Educational Research Association, 1972. (ERIC Document Reproduction Service No. ED 062 389)

Jones, H. L., Thompson, B., & Miller, A. H. How teachers perceive similarities and differences among various teaching models. Journal of Research in Science Teaching, 1980, 17, 321-326.

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

Kaiser, H. F., Hunka, S., & Bianchini, J. Relating factors between studies based upon different individuals. In H. J. Eysenck & S. B. G. Eysenck (Eds.), Personality structure and measurement. San Diego: Knapp, 1969.

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.

Miller, A. H., Thompson, B., & Frankiewicz, R. G. Attitudes of teacher education students toward teachers. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975. (ERIC Document Reproduction Service No. ED 104 887)

Morrow, J. P., & Frankiewicz, R. G. Strategies for the analysis of repeated and multiple measure 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.

Peterson, D. R. Scope and generality of verbally defined personality factors. Psychological Review, 1965, 72, 48-59.

Thompson, B. Teacher thought regarding instructional preferences. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, 1980. (ERIC Document Reproduction Service No. ED 196 918)

Thompson, B., Borgers, S., & Ward, R. G. Professional self-concept as a training construct for the helping professions. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. (ERIC Document Reproduction Service No. ED 156 977)

Thompson, B., & Miller, A. H. Dissonance theory and education students' attitudes toward teachers. Journal of Experimental Education, 1978, 47, 55-59.

Thompson, B., & Miller, A. H. The utility of "Social Attitudes" theory. Journal of Experimental Education, in press.

Thompson, B., & Pitts, M. C. Validity of teachers' perceptions of childrens' cognitive styles. Measurement and Evaluation in Guidance, in press.

Thompson, B., & Stapleton, J. C. A strategy for validating semantic differential referents. Journal of Experimental Education, 1979/80, 48, 110-113.

Table 1

Reliability Coefficients and Cosines

	2			5			9			17			33		
Factor	Alpha	I	II	Alpha	I	II	Alpha	I	II	Alpha	I	II	Alpha	I	II
I	.92	--	.81	.83	.81	-.70	.91	.93	.91	.96	.78	.99	.96	.99	--
II	.65	--	.96	.70	.91	.92	.61	.92	.78	.67	.91	.99	.67	.99	--
III	.56	--	.91	.60	.96	.83	.56	.83	.53	.50	.53	.99	.50	.99	--
IV	.39	--	.81	.53	.32	.95	.34	.97	.84	.44	.84	.99	.43	.99	--
V	.22	--	.32	.36	--	.99	.24	-.70	--						
VI	.08	--	-.44	.36	-.44	.76	.08	.95	--						
VII				.18	.82	.93	.01	.76	--						

<u>Note</u>. The columns headed "I" report the cosine between the factor for a row in the column and the related factor in the solution involving the next <u>smaller</u> number of scale steps. The columns headed "II" report the cosine between the factor for a row in the column and the related factor in the solution involving the next <u>larger</u> number of scale steps.