Reprinted from Multivariate Behavioral Research, July, 1974, Vol. 9, pp. 273-282

"SOMETIMES, IT'S OKAY TO FACTOR DIFFERENCE SCORES"—THE SEPARATION OF STATE AND TRAIT ANXIETY[1]

JOHN R. NESSELROADE[2]
Pennsylvania State University

and

DANA G. CABLE[2]
Hood College

ABSTRACT

Contemporary psychometric policy and practice have tended to make the use of algebraic difference scores in psychological research taboo. Within the more limited domain of factor analytic research on personality, difference scores have been the subject of sporadic debate for more than 30 years. Using the personality trait versus state distinction as a substantive context, the fit of the factor analytic model to difference score data is investigated and found to be quite good. Methodological issues related to properties of difference scores and their implications for personality research are briefly discussed.

The algebraic difference score, more specifically the raw gain score, is generally regarded by the psychometrically-oriented as one of the least respectable of quantitative indices despite its intuitive appeal in a variety of applications. Reaction to its use typically runs from indifference to active repugnance. In a recent paper by Cronbach and Furby (1970), the difference score is relegated to a local, if not an absolute minimum on the usefulness-popularity surface.

Within a factor analytic context, the difference score has a history of rejection beginning at least some 30 years ago (Woodrow, 1939). More recently it has been condemned by, for example, Harris (1963a), Horn (1963), and Humphreys (1961). Of the advocates of difference scores as proper data for factor analytic investigations, Cattell (1963, 1966), who uses difference score factoring (differential-*R* technique) to structure the multidimensional nature of change, has been the most outspoken.

In an attempt to set the emerging personality trait vs. state distinction (Singer & Singer, 1972) in a viable methodological framework, Nesselroade and Bartsch (1974) examined the algebraic relationships between factors obtained from separate analy-

1. Preparation of this paper was facilitated by U.S. Office of Education Research Grant O E G-3-71-0113. Requests for reprints should be mailed to: John R. Nesselroade, College of Human Development, Pennsylvania State University, University Park, Pennsylvania 16802.

2. Formerly at West Virginia University.

ses of pre- and post-test data and those obtained from analyzing the corresponding difference scores. Their general conclusion was that when traits and states were defined in terms of time-related properties of factor loadings and factor scores, the relations between separate occasion factors and difference score factors were predictable but, in some cases, at variance with projected relations based on alternative models.[3] The primary objective of the present report is to provide an empirical demonstration of these relationships within the anxiety state vs. trait context.

We note at the outset that there are troublesome scaling and estimation issues connected with certain uses of difference scores, but these will not be resolved here. The reliability of difference scores, a matter of general concern, will be discussed subsequently.

The development of multivariate models for investigating longitudinal change has received considerable attention in the past few years. In addition to the collection of papers edited by Harris (1963b), recent discussions by, for example, Corballis and Traub (1970), Hakstian (1971), and Horn (1963) have offered further solutions to the problem. Nesselroade and Bartsch (1974) considered four cases which, in terms of the factorial composition of the measures across time, were described as: (*a*) invariant loading pattern-stable factor scores, (*b*) invariant loading pattern-fluctuant factor scores, (*c*) non-invariant loading pattern-stable factor scores, and (*d*) non-invariant loading pattern-fluctuant factor scores. They concluded, for example, that separate occasion score matrices decomposable into factor solutions exhibiting stable factor scores (cases *a* and *c*) would yield a difference score solution whose loading pattern was the difference between the separate occasion loading patterns. Of course, if case *a* holds, since the factor loading patterns are invariant, the difference score factor loadings vanish altogether. The constraint of stable factor scores represents a trait model, and the conclusion reached by Nesselroade and Bartsch concerning relationships between difference and separate occasion loadings is quite in agreement with that of Harris (1963a) who develops a procedure for analyzing two occasion data which fixes factor scores to be perfectly stable across time.

When the separate occasion score matrices are not decomposable into factorial solutions exhibiting stable factor scores (cases *b* and *d*), a state model is implied and the conclusions concerning

3. In deriving the relationships described, an assumption made by Nesselroade and Bartsch was that the common factors on one occasion do not covary with the unique factors on the other occasion.

difference score factors are altered. For a discussion of case *d* the reader is referred to Nesselroade and Bartsch (1974). Case *b*, however, is germane for two reasons: (1) it has been neglected in critical discussions, although it is the most interesting from the standpoint of difference score factoring and (2) it provides an appropriate context within which to remark on the issue of reliability of difference scores.

If difference scores, derived from separate occasion score matrices describable as case *b*, are factored, the loading pattern will be proportional (by columns) to the (invariant) loading pattern of the separate occasion solutions. With proportionality of patterns established, one can, if one desires, rescale the separate occasion and difference score loading patterns to strict numerical invariance by adjusting the standard deviations of the corresponding factor scores. Thus, under case *b* conditions the difference score loadings are not the differences between corresponding separate occasion loadings but, rather, are identical to them.

But what of the notorious unreliability of differences scores? The writers believe that this question tends somewhat to resolve itself if placed in a trait vs. state perspective. Much of the earlier controversy over difference scores occurred within the context of ability and aptitude change wherein measures are often expected, and have been shown, to possess a substantial degree of stability (test-retest correlation). The classical test theory formulation of difference score reliability leads to discouragingly low reliability coefficients when test-retest coefficients are high. But this is precisely the case for which it generally makes little sense to compute difference scores. If, over time, a measure shows high stability with intra-individual change and the standard deviation remains relatively constant, the increments (or decrements) will tend to be fixed across individuals and the group mean suffices to describe change.

Consider, however, measures on which individuals tend to show quite differential amounts of change from one occasion to the next. If the reliabilities of the measures on the separate occasions are high, the lower the test-retest stability, the more acceptable the reliability of each of the difference measures. The writers are aware of that venerable specter (see, e.g., Bereiter, 1963)—"Are you measuring the same thing when test-retest stabilities are low?"—and propose that it can be temporarily exorcised within the content domain of psychological states, especially anxiety (Singer

& Singer, 1974), where conceptual formulation and empirical data agree that persons' scores change differentially over time on measures of high within-occasion reliability.

Thus a reasonable psychological situation—that factors may show invariance of loadings across occasions but individuals may show substantial differential change in their factor scores—may be postulated which produces data apparently quite amenable to difference score factoring. The significance is not that the difference score analysis yields factors that are derivatives of the separate occasion factors but, rather, that it yields the same factors; same in that the difference score factor pattern and the separate occasion factor patterns are invariant. Relationships among factor scores will be mentioned in the discussion section.

Design and Method

The present study was designed to examine empirically the substantively and algebraically expected relationships among factors obtained from separate occasion and difference score analyses. The choice of measurement variables was guided by an effort to mark two factors in the anxiety domain: (1) state anxiety—a dimension showing invariant loading patterns and differentially changing factor scores, and (2) trait anxiety—a dimension showing invariant loading patterns and stable factor scores.

In accord with the model discussed previously, if the expected factorial nature of the measurement variables obtains, the difference score analysis should produce one factor and the separate occasion analyses should each produce two factors. The difference score analysis factor loading pattern should match that of one factor from each occasion and the remaining separate occasion factor patterns should match each other. The former three factors would be expected to load saliently the state anxiety measures; the latter two factors to load the trait anxiety measures.

Measurement Variables

The State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1969) and an experimental set of state and trait anxiety scales currently under development by Cattell and Nesselroade (1974) were used as the measurement instruments. The Spielberger et al. questionnaire items were administered under the prescribed instructions ("Respond as you generally feel" for trait

items, "Respond as you feel right now" for state items). The Cattell-Nesselroade items are worded to include the sense of "right now" or "generally" and were all administered under standard questionnaire instructions.

The questionnaire scales were split (odd vs. even items) into eight subscale scores for each subject. This was done to provide an adequate set of market variables for hypothesized trait and state anxiety factors and also to provide within-occasion reliability estimates (split half). The eight subscales, their item composition, and estimated reliability coefficients (uncorrected for length of scale) are listed in Table 1.

Subjects and Procedures

One hundred forty-one undergraduates (90 males and 51 females) served as subjects in this research project. Subjects were randomly divided into three groups: (1) stress group, (2) relaxation group, and (3) control group. All subjects were administered the total set of questionnaire items twice, with an interval of 50 minutes elapsing between the two administrations. The purpose and nature of the treatments and the outcomes are described fully by Cable (1972). For present purposes, we merely note that the treatments provided a range of intra-individual changes from occasion 1 to occasion 2.

Table 1
Item Composition of the 8 Subscales of the Anxiety Battery and Reliability Coefficients

Variable	Items	Reliability Occasion 1	Reliability Occasion 2
1. State Anxiety	10 Spielberger State Items	.86	.87
2. State Anxiety	10 Spielberger State Items	.86	.87
3. State Anxiety	10 Cattell-Nesselroade State Items	.78	.82
4. State Anxiety	10 Cattell-Nesselroade State Items	.78	.82
5. Trait Anxiety	13 Cattell-Nesselroade Trait Items	.59	.61
6. Trait Anxiety	12 Cattell-Nesselroade Trait Items	.59	.61
7. Trait Anxiety	10 Spielberger Trait Items	.82	.85
8. Trait Anxiety	10 Spielberger Trait Items	.82	.85

Analysis and Results

A separate covariance matrix was obtained for each set of occasion data and for the difference scores. Each variable was arbitrarily scaled to unit variance over the pooled occasions of measure-

ment. The variance of each difference measure was then fixed, in accord with the formula for the variance of a linear combination, as the sum of the separate occasions score variances less twice their covariance.

Eigenvalues of the three covariance matrices were computed and inspected to determine the number of factors to be extracted from each. The eigenvalues are presented in Figure 1. As the reader can see, the distributions of eigenvalues rather clearly warrant extracting two factors from the separate occasion covariance matrices and one factor from that based on difference scores. A principal axes factor solution was subsequently obtained for each covariance matrix after iterating on the specified number of factors until communality estimates had converged through three decimal places.

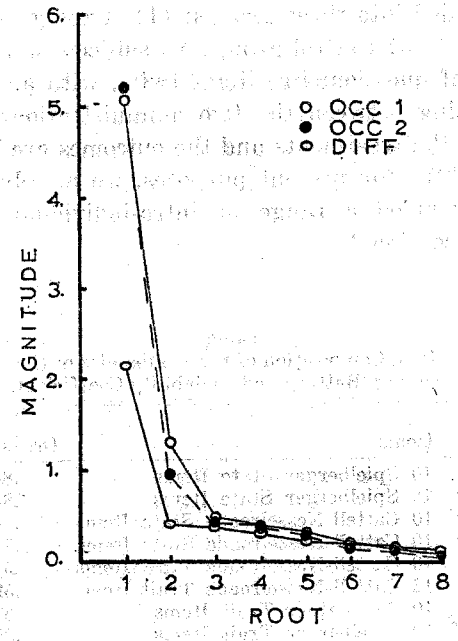

Fig. 1. Roots of Occasion 1, Occasion 2, and Difference Score Covariance Matrices.

Since the loadings of the single factor extracted from the difference score analysis were determined up to a constant of proportionality, the separate occasion solutions were rotated obliquely so that, for each occasion the loadings of one factor afforded a

least square fit to the loadings of the difference score factor. The second factor was targeted to load the four trait anxiety measure (variables 5, 6, 7, and 8). The rotated factor loadings and the factor covariance matrices are presented in Tables 2 and 3.

Table 2
Factor Loading Patterns

Variable	State Anxiety Occasion 1	State Anxiety Occasion 2	State Anxiety Differences	Trait Anxiety Occasion 1	Trait Anxiety Occasion 2
1. State Anxiety	.75	.63	.78	−.07	.05
2. State Anxiety	.76	.69	.78	−.11	−.05
3. State Anxiety	.53	.72	.57	.21	−.01
4. State Anxiety	.60	.62	.55	.09	.01
5. Trait Anxiety	.12	.06	.15	.65	.66
6. Trait Anxiety	−.01	.07	.15	.80	.71
7. Trait Anxiety	.20	.13	.10	.71	.79
8. Trait Anxiety	.18	.19	.16	.70	.75

Table 3
Factor Covariance Matrices

Factor	Occasion 1 I	Occasion 1 II	Occasion 2 I	Occasion 2 II
I	1.75	.67	1.61	.84
II	.67	1.00	.84	1.00

Difference score factor variance = 1.00

To summarize the degree of similarity among factor loading patterns, coefficients of congruence (Harman, 1960) were computed. These are presented in Table 4.

Table 4
Coefficients of Congruence Between Patterns

Occasion 2	Occasion 1 I	Occasion 1 II	Difference
I	.98	—	.98
II	—	.98	—
Difference	.99	—	—

Discussion and Conclusions

The results of the factor analyses are quite in line with both substantive and algebraic expectations. From a substantive viewpoint, the four variables included to mark state anxiety should cluster together on one factor, and the four trait anxiety marker

variables should cluster together on a second factor, as they do in the separate occasion solutions. Admittedly, the rotations leading to the separate occasion solutions were deliberately produced, but this can be justified for several reasons. Suffice it to say, for the present, that the clarity of the two-factor solutions is such that most oblique rotation procedures would produce highly similar configurations.

From an algebraic perspective, the following may be said: A factor, in this case trait anxiety, that shows an invariant loading pattern and highly stable factor scores from occasion 1 to occasion 2 should vanish when the difference scores are factored, as noted in the introduction. In contrast, a factor that has an invariant loading pattern but factor scores of lower stability, here state anxiety, should appear when the difference scores are factored. The obtained results fit these expectations rather closely.

The reader is reminded that the single principal axis extracted from the difference scores is not subject to further rotation. Its pattern, however, is rather clearly identifiable as state anxiety. In rotating the separate occasion solutions to produce similar patterns, the factor variances were allowed to depart from unity in order to hold the least squares fit. As reported in Table 3, state anxiety has a larger variance on occasions 1 and 2 than in the case of difference scores, but this is not unexpected. Had the model we are using fit the data exactly, the variance of the difference score factor would equal the sum of the variances less twice the covariance of the separate occasion state anxiety factors. Over 50 minutes the state measures exhibit positive stability. Under more extreme circumstances the state measures could show negative stability which would lead to a reversal of the variance relationships.

Interestingly, the separation of state and trait anxiety measures achieved here occurred without the two kinds of measures showing extremely disparate stability coefficients. The range of stability coefficients for the four state anxiety measures was +.57 to +.68 and that for the four trait anxiety measures was +.83 to +.87.

These data indicate a high degree of consistency of the factor model in relating separate occasion and difference score factoring in the anxiety questionnaire domain. Of course, the argument that one should know something about the substantive nature of his measures before factoring them cannot be overemphasized, and

this most assuredly holds for difference scores. However, from a structural point of view, difference scores offer one way of getting at the multidimensional nature of change. The data further show that, for some types of change, difference score factor analysis is not misleading but actually rather precise in not only revealing change dimensions but in suppressing dimensions of stable inter-individual differences.

References

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.) *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

Cable, D. G. Separation of state and trait anxiety in the context of short term change. Unpublished doctoral dissertation, West Virginia University, 1972.

Cattell, R. B. The structuring of change by *P*-technique and incremental *R*-technique. In C. W. Harris (Ed.) *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

Cattell, R. B. Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology.* Chicago: Rand McNally, 1966.

Cattell, R. B., & Nesselroade, J. R. *The state-trait anxiety battery (STAB).* Champaign, Illinois: Institute for Personality and Ability Testing, 1974.

Corballis, M. C., & Traub, R. E. Longitudinal factor analysis. *Psychometrika,* 1970, *35,* 79-98.

Cronbach, L. J., & Furby, L. How should we measure "change"—Or should we? *Psychological Bulletin,* 1970, *74,* 68-80.

Hakstian, A. R. Some notes on the factor analytic treatment of measures obtained on two different occasions. Research and Information Report (RIR-71-10), Division of Educational Research, University of Alberta, 1971.

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

Harris, C. W. Canonical factor models for the description of change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1963. (a)

Harris, C. W. (Ed.) *Problems in measuring change.* Madison: University of Wisconsin Press, 1963. (b)

Horn, J. L. The discovery of personality traits. *Journal of Educational Research,* 1963, *56,* 460-465.

Humphreys, L. G. Mimeographed paper on derived scores. Psychology Department, University of Illinois, 1961.

Nesselroade, J. R., & Bartsch, T. W. Multivariate experimental perspectives on the construct validity of the trait-state distinction. In R. B. Cattell & R. M. Dreger (Eds.) *Handbook of modern personality theory.* New York: Appleton Century Crofts, 1974. In press.

Singer, J. L., & Singer, D. G. Personality. In P. H. Mussen & M. R. Rosenzweig (Eds.) *Annual review of psychology.* Palo Alto: Annual Reviews, Inc., 1972.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. *The state-trait anxiety inventory (STAI) test manual: Form XX.* Palo Alto: Consulting Psychologists Press, 1969.

Woodrow, H. Factors in improvement with practice. *Journal of Psychology,* 1939, *7,* 55-70.