The clinical use of difference scores: Some psychometric problems

Item Type	Article
Authors	Cattell, Raymond B. (Raymond Bernard), 1905-1998
Citation	Cattell, Raymond B. (1982). The Clinical Use of Difference Scores: Some Psychometric Problems. Multivariate Experimental Clinical Research, 6(2), 87-98. https://doi.org/10.62704/10057/17540
DOI	10.62704/10057/17540
Publisher	Wichita State University, Department of Psychology
Download date	2026-05-19 19:00:14
Link to Item	https://soar.wichita.edu/handle/10057/17540

Multivariate Experimental Clinical Research, Volume 6, No. 2, 1982, 87-98

THE CLINICAL USE OF DIFFERENCE SCORES: SOME PSYCHOMETRIC PROBLEMS

Raymond B. Cattell
Hawaii School of Professional Psychology

ABSTRACT

Some pure statisticians have raised such difficulties regarding difference scores that clinicians and others who are bound to use them have been doing so only with trepidation.

This article examines the dependability of difference scores as a function of (1) the dependability of the single occasion scores, (2) the correlation between pre- and post-interval scores, (3) the pre-post difference of means and (4) the relative pre-post standard deviation —all in regard to the same variable measured twice on the same people.

Uncorrelated pre- and post-scores, equal or unequal in variance, yield differences with no larger percent error than the single occasion scores. The difference score has highest dependability with a negative pre-post correlation, and is reduced finally to zero with a positive, except when before and after standard deviations are different.

The suitability of procedures depends on the psychological model one is using — merely itemetric or heeding structures recognized in trait and state theory. In the latter one must distinguish between the "instant" pre-post *dependability* coefficient and the long term *stability* coefficient, in which the true score itself alters. A *trait constancy* coefficient of 0.5, not uncommon over, say, six months of therapy, results in a reduction of a difference score dependability coefficient of 0.9 to 0.82, which can readily be compensated by a Spearman-Brown calculated increase of test length. More important than what some statisticians have emphasized is the need for getting equal interval properties in the test, by pan-normalization or relational simplex principles.

INTRODUCTION

Clinical psychology seems long to have been content to use measurement only in initial personality assessment and diagnosis. With the advent of more analytic measures of factor source traits, as in the 16 PF, the Guilford-Zimmerman, the HSPQ, MAT, etc., it is now possible to monitor progress and adjust therapeutic tactics to the changes found. A rationale for such procedures, with illustrations, has recently been given by Cattell (1980), combining measurement with the principles of structured learning theory.

In the end, all such clinical direction hinges on the reliability of change measures, i.e. of a difference score on the same trait or state measure between two occasions on the same person. The provocation of the present article is that certain statisticians, e.g. Cronbach & Furby (1970) started such doubts about difference scores that even committees for graduate students' theses have been known to deplore the use of difference scores — and this despite the adequate rebuttal by Nesselroade & Cable (1974), and the prior analysis (1966) by the present writer. Indeed, as recently as 1981, a research proposal by a colleague was rejected by a Washington committee, inaptly and ineptly repeating the old criticism.

Actually, some essentials in the use of difference scores had been set out way back by Lord (1958, 1963), the present writer (1966), Damarin, Tucker & Messick (1966) and others. The issues arise both in ANOVA and CORAN (correlation analysis designs) and have been sufficiently dealt with in the former in the principles for repeated measurements designs. It is in CORAN that the problems arise that we shall presently discuss.

These problems concern the basic approaches to trait and state definition. The discovery of the main source traits in our culture has rested on R-technique factor analysis, i.e. factoring one-occasion measures. Test consistency and validity for such structures are commonly understood. State and trait-change factors, on the other hand, depend for their recognition on dR (differential R) technique by factoring difference scores — over a few days for the former and months or years for the latter. The relation of trait (R-technique) factors to trait change (dR-technique) factors is particularly instructive for developmental psychology (Nesselroade & Reese, 1970). This is one of several areas where the insightful handling of difference scores is important.

INFLUENCES AFFECTING DIFFERENCE SCORES

What has upset people about difference scores is that at first glance they would appear to have twice the proportion of error variance of single scores, because there is random error in the first measurement and again in the second — a double error contribution as it were.

But let us pursue a model in which there is a fixed trait score t (t_1 or t_2) on both occasions. (This could be zero in a state). There will next be a fluctuation from that score, t_f (t_{f1} and t_{f2}) present on both occasions. (This is all that will be real in a true state.) Finally there will be sheer random measurement error e (e_1 and e_2).

The magnitude of the observed score o (o_1 and o_2) on one occasion will be

$$o_1 = t_1 + t_{f1} + e_1 \tag{1}$$

and similarly for o_2.

The difference score will be

$$o_{(2\text{-}1)} = (t_2 - t_1) + (t_{f2} - t_{f1}) + (e_2 - e_1) \tag{2(a)}$$

which, since $t_2 = t_1$ will be

$$o_{(2\text{-}1)} = (t_{f2} - t_{f1}) + (e_2 - e_1) \tag{2(b)}$$

The variance of this difference would contain negative covariances if any correlation among $t_1, t_2, t_{f1}, t_{f2}, e_1$ and e_2 exists but none does. (If derived from 2(a) we should have $r_{t_1 t_2} = -1.0$.)

So we have:

$$\sigma^2_{o(2-1)} = 2\sigma^2_{tf1} + 2\sigma^2_{e1} \qquad (3)$$

If instead of separating the trait part into a fixed and fluctuant part we had left it as t'_1 ($= t_1 + t_{f1}$) then $t_1{}'$ and t'_2 would correlate positively and the whole observed score of the two measures involved in the difference score would do so. The doubters of difference scores, using that model, rightly perceive that the ensuing positive correlation will produce a negative covariance, so that the ratio of true to error score in the difference score will be lower than in the single occasion scores. However, we wish here to see how things will behave in what seems a more realistic model.

Historically, let us note that this tendency for a significant positive correlation between pre- and post-experimental (or developmental) scores led to Lord (1958) and other competent statisticians in psychometry, to suggest using difference scores derived by partialling out the pre from the post score, i.e. removing the correlation. By some this was viewed also as taking the error on occasion 1 out of occasion 2 and thus not dealing with double error. (Some ingenious variations partialled the *mean* of pre- and post- from the difference.) This has a certain neatness, but it is an arbitrary emasculation of nature that would remove all chance of discovering any natural laws that might exist connecting original and difference scored. For example, there are growth curves where growth is a function of the absolute level, and homeostatic laws where change is a function of initial deviation.1

Our position will be that to tamper with the raw difference score (other than the usual use of standard score transformations) is to ignore the most likely model and certain real relations to it. If the proof of the pudding is the eating then the internal and external consistency of the scientific results we have so far obtained in factoring difference scores (dR technique) in programmatic work (Cattell, 1979-1980; Horn, 1972) supports the direct — one might say honest — use of untouched raw score differences. Admittedly particular precautions need to be taken in so doing to reduce the admitted effect of pre-post correlation. These precautions have been described elsewhere and include (1) Using longer tests to raise the dependability coefficient of the single occasion measures themselves, (2) Keeping conditions such as to maintain pre- and post- dependabilities essentially equal, (3) Keeping pre- and post- variances equal, if necessary by putting them in standard scores, and (4) Obtaining both *dependability* (re-test without time for trait change) coefficients and *stability* (re-test after psychological change) cofficients, to use in ways described below.

The precautions thus seen as necessary can be considered dictated by the following causes of blurring of difference factor patterns (See Bartsch & Nesselroade, 1973). In summary they are:

(1) That since raw scores from scales are not guaranteed to have equal interval properties this will bring more serious distortion to difference scores than to single occasion scores (Cattell, 1966, p. 366; 1978, p. 345).

(2) That if there is a large positive correlation between the before and the after scores "the amount of error in a difference score can be about double that in a

single score" (Cattell, 1973). In Table 1 this is illustrated and supported in equation 4 by taking a fairly usual true score correlation of 0.5 between occasions (See Schuerger, Tait & Tavernelli, 1982). Naturally the above statement would be misleading if less usual correlations were encountered.

(3) That the problem is increased by low reliability of the single occasion scores themselves.

(4) That false factors (traits, not states) will intrude if there are gross difference of raw score *standard deviation* between the two occasions.

ESTIMATION OF THE EFFECT OF TWO MAIN INFLUENCES ON THE DEPENDABILITY COEFFICIENT

We set the first and fourth influence above aside, since the latter is likely to stay fixed in most clinical studies (but perhaps not in learning experiments) and we shall soon suggest ways of reducing the former. Accordingly we shall now evaluate the psychometric impact of the second and third, as they apply to difference scores (in ANOVA or CORAN) in use against other, dependent or independent variables.

The discussion will go beyond that commonly presented in bare operational pyschometric coefficients in as much as we shall consider application to models of *traits, states* and *trait development factors.* The reader would do well to refresh himself on more recent (Cattell, 1973, 1978) analytical concepts concerning coefficients of test consistency and validity. For it is necessary to recognize the distinction of *dependability, homogeneity,* and *transferability,* as distinct aspects of *test consistency.*[1] Note also that the *constancy coefficient,* r_c, applies to a *trait,* not a test though it is derived from test-operational *dependability* r_d and *stability,* r_s coefficients.

Concerning the relation of correlation to variance pedicted it should be recognized that in all re-tests, i.e. with the same test, the coefficient *directly* represents the percent of shared variance. (The *squared r,* as usual, gives prediction of variance of one testing accounted for by the totality of the other.) Thus an r_d (dependability) of 0.85 indicated 85% of true, common variance in each and 15% of error or non-common variance.[2]

It is necessary to set out the effect of the components in equation (2(b)) upon the test variance. Since these components are uncorrelated, the deviation on occasion 2 having nothing to do with that on 1, the difference score variance will be:

$$\sigma^2_{o(2-1)} = \sigma^2_{tf1} + \sigma^2_{tf2} + \sigma^2_{e1} + \sigma^2_{e2} \tag{4}$$

which, with assumption of equal variances on the two occasions we reduced to (3) earlier. The dependability coefficient for the difference score will thus be:

$$r_{d(2-1)} = \frac{\sigma^2_{tf}}{\sigma^2_{tf} + \sigma^2_{e}} \tag{5}$$

This could be directly observed and calculated only if we could *exactly* repeat the experiment and the subjects' states in coming to the second difference experiment and our concern is therefore to *calculate* (5) from other sources. Those

sources, which we agree are the dependability of the single occasion scores and the correlation of pre- and post-scores are represented in variance terms as follows:

Dependability (Immediate retest)

$$r_{d(t1)} = \frac{\sigma^2_{t1} + \sigma^2_{tf1}}{\sigma^2_{t1} + \sigma^2_{tf1} + \sigma^2_{e1}} \tag{6}$$

and:

Stability (Long term retest)

$$r_{s(t1)} = \frac{\sigma^2 t1}{\sigma^2 t1 + \sigma^2 tf + \sigma^{2'} e} \tag{7}$$

The coefficient of trait (or state) constancy (which is *not* a property of the test) is:

$$r_{c(t1)} = \frac{\sigma^2 t1}{\sigma^2 t1 + \sigma^2 tf1} \tag{8}$$

The algebra of deriving $r_{d(2\text{-}1)}$ from these involves several steps which can be sketched in the main links:

$$\frac{1}{r_d - r_s} = \frac{\sigma^2 t + \sigma^2 tf + \sigma^2 t}{\sigma^2 tf} \tag{7}$$

$$\text{(Substituting } r_{d(2-1)} = \frac{1}{r_{d(2-1)}} + \frac{\sigma^2 t}{\sigma^2 tf} \tag{8}$$

$$\frac{\sigma^2 t}{\sigma^2 tf} = \frac{r_c}{1 - r_c} \tag{9}$$

(Substituting in (8))

$$r_{d(2-1)} = \frac{r_d - r_s}{\frac{r_c}{1 - r_c}(r_d - r_s)} \tag{10}$$

This is the required derivation of the directly unobservable $r_{d(2\text{-}1)}$ from properties of the single occasion scores. But actually r_c can scarcely be called directly observable and fortunately we know (Cattell, 1973) that:

$$r_c = \frac{r_s}{r_c} \tag{11}$$

which nicely simplifies (10) to

$$r_{d(2-1)} = \frac{r_d - r_s}{1 - r_s} \tag{12}$$

To show what this means in terms of a frequently encountered range of values we have taken in Table 1 dependability coefficients of .9, .8, and .7, and stability coefficients of .7, .6, .5, 0 and -.5.

TABLE 1

DEPENDABILITY COEFFICIENTS OF DIFFERENCE SCORES, $r_{d(2\text{-}1)}$, DERIVED FROM VALUES OF SINGLE OCCASION SCALES

		Dependabiltiy Coefficient of Single Scale		
Stability Coefficient by Single Scale		.90	.18	.70
	.7	.66	.33	.0
	.6	.75	.50	.25
	.5	.80	.60	.40
	0	.90	.80	.70
	-.5	.93	.87	.80

It will be seen at once that with commonly encountered values — the stability coefficient over several months of about 0.5 (Schuerger, Tait & Tavernelli, 1982) or a couple of years of about .3 (Cattell and Scheier, 1961)[3] — the reliability of a difference score is quite substantial. Much research has to be satisfied with this dependability of 0.6 to 0.8 even in single occasion scores.

When the stability coefficient rises to equal the dependability we see (top right Table 1) that difference scores cease to have any dependability. This means, in variance terms, that the fixed trait score remains fixed and that the t_f term is zero. All beyond the fixed trait is error of measurement and the difference is only error. A model with only *t* and *e* ignoring the difference of the *dependability* and *stability* consistency coefficients, has led some analysts to conclude that this is the nature of *any* difference score. Consequently the first *dR* technique factor analyses were greeted with the criticism that one was only factoring error.

An interesting case occurs when the stability is zero. This means, in variance terms, that there is no *t*, fixed trait value, and the constancy coefficient (Equation 11) is also zero, no matter what the test dependability. Here we have evidence that we are dealing with a pure state measure. Incidentally this should bring home the conceptual error of considering r_c and r_s as properties of a *test* rather than a *trait*, and the foolishness of the terms non-concurrent validity and concurrent validity. The essential conclusion here (row 4 of Table 1) is that when we are dealing with state measures, establshed by *dR*- or P-technique (Curan & Cattell, 1978) the dependability of a difference score equals that of the original testing instrument. (This can also be seen from the variance of the difference having the same $\sigma^2{}_t / \sigma^2{}_e$ ratio, as shown in Equation 3.)

Experimental situations may arise in which the stability is actually negative. If "the meek shall inherit the earth" the meeker priest might become, by selection, the prouder prelate. Or if a group of cyclic manic patients were tested in the manic phase it could be that at a six-month retest the clearest cyclic cases would show the deeper depressions. Whatever the experimental or developmental causes, a negative stability coefficient, as illustrated by -.5 in Table 1, will result in the dependability of the difference score *actually exceeding that of the original single occasion test.*

So far we have proceeded with the usually reasonable assumption that the variance components, $\sigma 2_t$, $\sigma 2_{tf}$, and $\sigma 2_c$ will be essentially the same in proceeding from test to re-test. However it *might* alter, and, indeed, in the case of a group-measured on, say, anxiety and re-measured under the influence of a powerful common anxiety stimulus, modulation theory expects the standard deviation to change proportionally to the change in mean (Cattell, 1979). Indeed, Cattell & Brennan (In press) have shown that in both anxiety and depression the group variance increases proportionally to the increase of the group mean through the anxiety or depression provoking modulating stimulus situation. It seems best to handle this differently from Table 1, namely by showing the variance changes which bring about the stability and constancy coefficient changes. For simplicity we have held to a first administration dependability of 0.9 in column 1 Table 1. By modulation theory (Cattell, 1973, 1979; Cattell & Brennan, In press)

1st Occasion $S_{ik1} = s_{k1}L_i$ (13(a))
2nd Occasion $S_{ik2} = s_{k2}L_i$ (13(b))

where S_{ik} is the state score of person i in situation k, L_i is his prneness to that emotion, and s_k is the modulator strength of the stimulus situation k. In our notation so far S_{k1} will equal $(t_1 + t_{f1})$ and S_{k2} will equal $(t_1 + t_{f2})$ so we can let $L = t$ and write

$$t_1 + t_{f1} = s_{k1}t_1 \tag{14(a)}$$

and

$$t_1 + t_{f2} = s_{k2}t_1 \tag{14(b)}$$

whence $t_f = (1 + s_k)t_1$ in general.

To shift from modulator language to our present, let us call $(1 + s_k)$ some single multiplier m. Then on the first occasion of measurement

$$o_1 = t_1 + m_1t_1 + e_1 \tag{15(a)}$$

and on second

$$o_2 = t_1 + m_2t_1 + e_2 \tag{15(b)}$$

The difference score will be

$$(o_2 - o_1) = (m_2 - m_1)t_1 + (e_2 - e_1)$$

and the variance

$$\sigma^2 o(2\text{-}1) = \sigma^2(m_2 - m_1)t_1 + 2\,\sigma^2_e$$

extending the first term on the right we have

$$\sigma^2_{o(2\text{-}1)} = \sigma^2_{m_2t_1} + \sigma^2_{m_1t} - 2r_{m_2t_1m_1t_1}\,\sigma_{m_1t_1}\,\sigma_{m_2t_2} + 2\,\sigma^2_e$$

r is here the constancy coefficient of the trait. If it were unity, i.e. if no influence but the s_{k_2}-s_{k_1} difference affected all individuals between the two occasions then the variance would be

$$\sigma^2_{o(2\text{-}1)} = (\sigma_{m_2t} - \sigma_{m_1t})^2 + 2\,\sigma^2_e$$

and the dependability would be

$$r_{d(2\text{-}1)} = \frac{(\sigma_{m_2t} - \sigma_{m_1})^2}{(\sigma_{m_2t} - \sigma_{m_1t})^2 + 2\,\sigma^2_e}$$

In short, even with perfect trait constancy the difference score would have some dependability if the standard deviation altered between the two occasions. If influences additional to the experimentally imposed s_{k_1} and s_{k_2} modulated the liability, and were peculiar to each individual, $r_{m_2t.m_1t}$ would be less than unity and $r_{d(2\text{-}1)}$ would be larger.

A word is necessary, finally, about the apparent determination of the difference score dependability, $r_{d(2\text{-}1)}$, in equation (10) by all *three* coefficients, r_d, r_s, and r_c. As equation (12) shows there are actually only two, due to the interdependence in the three (Equation 11). Theoretically it is more basic to think of determination by the inherent constancy over time of the trait, r_c, and the sheer dependability of the test being used, r_d. Practically the values one will want for solution are an immediate test-retest, r_d, and a retest over the period in which one is interested in change from an experimental influence or life development.

SUMMARY

(1) Rumors of statistical condemnation of difference scores as excessively unreliable have to some extent inhibited their use where closer analysis would show it is appropriate.

(2) Bare discussion of coefficients without an underlying model has been partly responsible for uncertainties. Here the same basic model is proposed as has been effective on other areas. It rests on the findings by *R*-, *dR*- and *P*-technique factor analysis, in personality and ability modalities, of *fixed traits, trait change patterns,* and *state change dimensions.* A test score on one occasion can be divided into components *t* (fixed trait), *tf* (function fluctuation, either from trait or state change source) and *e* error of measurement.

(3) The coefficients affected by these components are studied in the light of the more recent treatment of *consistency* of a scale (Catttell, 1972), analyzing it into aspects of (a) *reliability,* (b) *homogeneity,* and (3) *transferability,* and operationally further breaking reliability into *dependability* and *stability* coefficient forms.

(4) Four main influences affect the dependability coefficient of a difference score (i) absence of equal interval properties on the single occasion scale reduce it, (ii) positive *stability* coefficients between test and retest, due to high *trait*

constancy coefficients (r_c's) reduce it, (iii) low *dependability* of the single occasion scales reduce it, and (iv) gross differences of standard deviation between the two occasions will reduce it.

(5) This article concentrates on (ii) and (iii) and shows, by analyzing the variances entering into the various coefficients that (with assumption of stability of each variance across two similar testings) the dependability of a difference score is calculable as:

$$r_{d(2-1)} = \frac{r_d - r_s}{1-r_s}$$

the dependabilities and stabilities being those of the single occasion.

Alternatively this can be stated in terms of the dependability and the trait constancy by

$$r_{d(2-1)} = \frac{r_d(1-r_c)}{1-r_c \cdot r_d}$$

the first deriving from observable coefficients, the second from personality theory concerning trait constancy.

(6) A table sets out the $r_{d(2-1)}$ values from commonly encountered r_d's and r_s's, in which it appears that (i) there is no difference score dependability when $r_d = r_s$, (ii) difference score dependabilities are quite good with r_d's about .8 to .9 and r_s's about 0 to .7, (iii) with zero stability, which is the perfect *state* scale finding, the difference scores have the same reliability as the single measure, (iv) in the unusual event of negative stability coefficients the difference score exceeds in dependability the single score.

(7) Modulation theory suggests that in state measurement the assumption of constant variance at different situational modulator indices cannot be made. As far as pursued here this seems not to invalidate the general evaluation of difference dependabilities, but it does show that with perfect trait constancy and high stability coefficients, that would on the first assumption reduce difference dependability to zero, the differences now retain dependability.

(8) The importance of change measurement has long been recognized in experiment, though more in bivariate ANOVA, than, until lately, multivariate CORAN designs. Its importance in clinical psychology has only appeared recently with the advent of tests capable of *meaningfully* measuring therapeutic change (Karoly & Steffen, 1980; Waskow & Parloff, 1975; Cattell, Rickels, et al. 1966; Rickels & Cattell, 1969). The studies and experiments, however, have so far scarcely once given attention to the basic problem of dependability of ordinary difference scores.

(9) As change measurement comes to play a more important role it becomes vital (i) to evaluate in each experiment the dependability ascribable to the given difference scores, (ii) to give fuller meaning to the measures in terms of (a) existing knowledge of traits and states (rather than mere variables) involved, (iii) to be prepared to give more testing time, in order (by the Spearman-Brown formula), to raise the dependability of the single measures to a level needed, in relation to the anticipated stability coefficients, (iv) to give more serious attention to the importance of equal unit scales in these areas. There are already available two approaches — *pan-normalization* (Cattell, 1973) and the *relational simplex*

(Cattell, 1962, 1973) — that could improve our scales in that respect.

(10) Difference measurement needs bringing into relation not only with factorial evidence on traits, trait change patterns and state dimensions, as suggested above, but also to models separating genetic maturation and structured learning (Baltes & Nesselroade, 1973; Cattell, 1982; Goulet, 1973; Schaie 1973) and such analyses of state and situation interaction as modulation theory (Cattell & Brennan, In press).

FOOTNOTES

[1] It will be noted that we keep here to the more analytical and precise terminology in the field of *test consistency* used elsewhere by the present writer (1973, p. 354) his coworkers and others, namely:

r_d = *dependablity* (short-term inter occasion, repeat reliability)

r_s = *stability*, repeat dependability over long term and situation change

r_{tc} = *trait constancy*, constancy of the trait itself (true score) over long term, situational change. Alternatively written $r_{at.bt}$ here.

[2] When one is involved in testing unitary source traits one must note that common variance need not all be the *wanted* factor of one's research, but is both the wanted and the unwanted broad or specific factor variance, i.e. "true" does not mean "pure factor" (See diagram, Cattell, 1973, p. 380).

[3] Some trait stability coefficient over truly long periods were provided by Kelley (1955) and over fairly long time intervals, as by-products, in the work of Cattell & Scheier (1961). The latter found for the pure anxiety factor over 3 years, for a group of medical students, a stability coefficient of 0.4. Many other such results on factored source traits must now be available for collation from the literature. A valuable comprehensive survey averaging personality trait test stability coefficients, by Schuerger, Tait, & Tavernelli (1982) points to 0.8 over a month, and about 0.6 over four years. With the exception of intelligence, and for such periods as are covered in developmental studies, an r_s of 0.6 such as we have taken for calculation is pretty representative.

REFERENCES

Baltes, P. B. & Nesselroade, J. R. The developmental analysis of individual differences on multiple measures. Pp. 219-252 in Nesselroade, J. R., & Reese, H. W. (Eds.) *Life span developmental psychology*, New York: Academic Press, 1973.

Bartsch, T. W., & Nesselroade, J. R. Test of the trait-state distinction using a manipulative factor analytic design. *Journal of Personality and Social Psychology*, 1973, 27, 58.

Cattell, R. B. The relational simplex theory of equal interval and absolute scaling. *Acta Psychologica*, 1962, 20, 139-158.

Cattell, R. B. Patterns of change: measurement in relation to state dimensions, trait change, liability and process concepts. In Cattell, R. B. (Ed.) *Handbook of multivariate experimental psychology,* Pp. 355-402. Chicago: Rand McNally, 1966.

Cattell, R. B. Evaluating therapy as total personality change: theory and available instruments. *American Journal of Psycotherapy,* 1966, 20, 69-88.

Cattell, R. B. Estimating modulator indices and state liabilities. *Multivariate Behavioral Research,* 1971, 6, 7-33.

Cattell, R. B. *Personality and mood by questionnaire.* San Francisco: Jossey Bass, 1973.

Cattell, R. B. Real base true zero factor analysis. MBR Monograph. Fort Worth: Texas Christian Univ. Press, 1972.

Cattell, R. B. *The scientific use of factor analysis.* New York: Plenum, 1978.

Cattell, R. B. The structured learning analysis of therapeutic change and maintenance. Pp. 163-193 in Karoly, P., & Steffens, J. J. (Eds.) *Improving the long term effects of therapy.* New York: Gardner Press, 1980.

Cattell, R. B. *Personality and learning theory.* Vols. 1 and 2. New York: Springer, 1979-80.

Cattell, R. B. & Brennan, J. State measurement: a check on the fit of the modulation theory model to anxiety of depression states. (In press, 1982)

Cattell, R. B. & Brennan, J. A computer program for achieving equal interval scales by the pan-normalization principle. (In preparation.)

Cattell, R. B., Rickels, K., Weise, C., Gray, B., & Yee, R. The effects of psychotherapy upon measured anxiety and regression. *American Journal of Psychotherapy,* 1966, 20, 261-269.

Cattell, R. B. & Scheier, J.H. *The meaning and measurement of neuroticism and anxiety.* New York: Ronald press, 1961.

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

Curran, J. P., & Cattell, R. B. *The IPAT 8-state Questionnaire.* Champaign, Ill.: IPAT, 1976.

Damarin, F., Tucker, L. R., & Messick, J. A base-free measure of change. *Psychometrika,* 1966, 31, 457-473.

Goulet, L. R. The interace of acquisition: models and methods for the developing organism. In Nesselroade, J. R., & Reese, H. W. *Life span developmental psychology.* New York: Academic Press, 1973.

Horn, J. L. State, trait and change dimensions of intelligence. *British Journal of Educational Psychology,* 1972, 42, 159-185.

Karoly, P., & Steffens, J. J. *Improving the long term effects of psychotherapy.* New York: Gardner Press, 1980.

Lord, F. M. Further problems in the measurement of growth. *Educational and Psychological Measurement,* 1958, 18, 437-454.

Lord, F. M. Elementary models for measuring change. In Harris, C. W. (Ed.) *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

Kelley, E. L. Consistency of the adult personality. *American Psychologist,* 1955, 10, 659-681.

Nesselroade, J. R., & Cable, D. G. Sometimes it's O.K. to factor difference scores —the separation of state and trait anxiety. Multivariate Behavior Research, 1974, 9, 273-282.

Nesselroade, J. R. & Reese, H. W. *Life span developmental psychology: methodological issues.* New York: Academic Press, 1973.

Rickels, K., & Cattell, R. B. The relation of clinical symptoms and IPAT factored tests of anxiety, regression and asthenia. *Journal of Nervous and Mental Disease,* 1968, 146, 147-160.

Schaie, K. W. Methodological problems in descriptive developmental research on adulthood and aging. Pp. 253-280 in Nesselroade, J. R. & Reese, H. W. (Eds.) *Life span developmental psychology.* New York: Academic Press, 1973.

Schuerger, J. M., Tait, E., & Tavernelli M. Temporal stability of personality by questionnaire. *Journal of Personality and Social Psychology,* 1982 (In press).

Waskow, J. E., & Pavloff, M. B. (Eds.) *Psychotherapy change measures.* Washington, D.C.: National Institute of Mental Health, U.S. Government Printing Office, 1975.