PSYCHOMETRIKA—VOL. 31, NO. 4
DECEMBER, 1966

A BASE-FREE MEASURE OF CHANGE*

LEDYARD R TUCKER

UNIVERSITY OF ILLINOIS

AND

FRED DAMARIN AND SAMUEL MESSICK

EDUCATIONAL TESTING SERVICE

A model for the measurement of the discrepancy between two scores is presented and discussed as a paradigm for the study of growth or experimentally produced change. The model assumes two tests or measures differing in complexity, and it analyzes the true difference between the test scores into a component that is entirely dependent on the first or base-line test and a second component that is entirely independent of it. Equations for estimating both components are given and these are compared with other measurement efforts with similar goals.

When two different tests have been adminstered to the same group, psychologists and educators often wish to summarize relationships between the two scores by means of a single number. This number, for example, sometimes represents readings from two different instruments given on a single occasion, as when performance on the verbal and quantitative subscores of the Wechsler Intelligence Scale are compared. Sometimes the number represents a discrepancy between scores on the same instrument given on two different occasions, as when tests are used to measure maturation, learning, or experimentally produced behavior change. For these purposes ratio, difference, and residual scores have been advocated and used in computing F tests, t tests, correlations, etc. This paper will derive and discuss measures of change or discrepancy that are primarily intended for correlational work. These measures will be called "independent and dependent change scores" even though they may be used to compare two different sorts of measures given simultaneously as well as the same measure given on two different occasions. We will compare the logic of these change scores with the logic of residual scores and difference scores and will offer some suggestions

*This research was supported in part by a grant from the United States Public Health Service, National Institute of Mental Health No. M-4186. We are greatly indebted to Harold Gulliksen and Frederic Lord for a careful reading of this manuscript and for their many valuable suggestions. The work of Gerri Nagy and Marie Davis in typing and proofreading is gratefully acknowledged.

about which of these is most likely to be appropriate to different correlational research problems.

A major attribute of data to be considered is the lack of perfect reliability. The basic data for both the residual and difference measures are fallible scores from ratings, naturalistic observations, tests, or other instruments. This fallibility of the basic measures leads to problems in the use of difference measures and to questions about the appropriate form and properties of residual scores. Some of the points to be considered here relate to the reliability of derived measures such as difference scores and residual scores. However, the more important issues are the underlying nature of the derived measures. While effects of unreliability of a derived measure on correlations of this measure with other variables may be accounted for by corrections for attenuation, the effects of automatic shifts in the underlying variable, or true score, induced by unreliability of the basic data can not be accounted for by subsequent corrections for attenuation. The unreliability of the basic data must be taken fully into account in the formulation of the derived measures themselves.

The raw score on measure X can be subscripted to indicate (i) the person to whom it applies and (ii) the occasion on which it was made. Thus, X_{ik} represents an observed test score (or other measurement) of individual i ($i = 1, 2, \cdots, N$) on test X on occasion k ($k = 1, 2$).

The notation for population values rather than sample statistics will be used to develop equations for several sorts of scores, but formulas for practical work will usually be given both ways. For covenience in deriving correlational formulas we will use deviation scores rather than raw scores. Thus, $x_{ik} = X_{ik} - E(X_k)$, where $E(X_k)$ is the expected value or population mean of the test X on occasion k.

Independent Gain Scores Compared with Difference Scores

Following the usual assumptions of test theory (Gulliksen [3]) we will represent observed deviation scores on the test X as the sum of true score and error components

(1) $$x_{i1} = t_{xi1} + e_{xi1} ,$$

(2) $$x_{i2} = t_{xi2} + e_{xi2} .$$

Some Properties of Difference Scores

The most obvious way to represent the discrepancy between two scores is by means of a raw difference score, which can be denoted by d_{xi} ,

(3) $$d_{xi} = x_{i2} - x_{i1} .$$

On substituting (1) and (2) in (3), we have

$$d_{xi} = (t_{xi2} - t_{xi1}) + (e_{xi2} - e_{xi1}). \tag{4}$$

The difference between the true scores is the true difference, denoted as δ_{xi} .

$$\delta_{xi} = t_{xi2} - t_{xi1} . \tag{5}$$

Substituting (5) into (4)

$$d_{xi} = \delta_{xi} + (e_{xi2} - e_{xi1}). \tag{6}$$

The observed difference score can be seen to consist of true score and an error component like any ordinary test score.

Some Properties of True Independent and Dependent Change Scores

These preliminaries lead directly into the "true independent change score" which is a major topic of this paper. The true score on the second, more complex test, t_{xi2} , can be defined as the sum of two independent components, one of which is specific to this second test and one of which is wholly predictable from the true score on the first test. The first or specific component is the true independent gains score, written γ_{xi} . These assumptions lead to the equation

$$t_{xi2} = a_x t_{xi1} + \gamma_{xi} . \tag{7}$$

The a_x in (7) is the coefficient for the regression of the true scores from the second testing on the true scores from the first testing. By revising (7) the true independent change score becomes

$$\gamma_{xi} = t_{xi2} - a_x t_{xi1} , \tag{8}$$

and the relationship between this new score and the more conventional difference score is seen to be

$$\delta_{xi} = \gamma_{xi} + (a_x - 1) t_{xi1} . \tag{9}$$

Equation (9) says that the true difference score can be expressed as the sum of two components: the true independent gains score, γ_{xi} , and a second component which is entirely dependent on the first test. We may call this latter component a true *dependent* change score and designate it ζ_{xi} .

$$\zeta_{xi} = (a_x - 1) t_{xi1} . \tag{10}$$

Therefore,

$$\delta_{xi} = \gamma_{xi} + \zeta_{xi} , \tag{11}$$

and the true difference score will equal the true independent change score only in the special case where $a = 1$ and $\zeta_{xi} = 0$.

When scores are defined in this way, the correlation between the true independent change score, γ_x , and the true score for the first test is identical

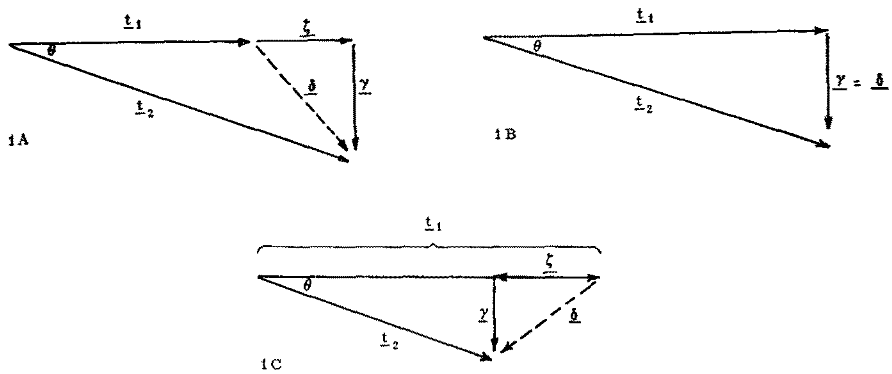

FIGURE 1
Relationships in terms of vector diagrams.
In Fig. 1A, $a > 1$;
in Fig. 1B, $a = 1$;
in Fig. 1C, $a < 1$.

to zero, as is the correlation between the true independent change score, γ_x , and the true dependent change score, ζ_x . Thus,

$$\rho_{\gamma_x t_{x1}} \equiv \rho_{\gamma_x \zeta_x} \equiv 0. \tag{12}$$

As a consequence, the variance in the true difference scores is the sum of the variances of the true independent change score and true dependent change score,

$$\sigma^2_{\delta_x} = \sigma^2_{\gamma_x} + \sigma^2_{\zeta_x} . \tag{13}$$

The relationships among the variables t_{x1} , t_{x2} , δ_x , γ_x , and ζ_x can be represented by the vector diagrams shown in Fig. 1 if the length of each vector is set equal to the standard deviation of the corresponding true scores. In these diagrams the x subscript has been dropped. Each of the diagrams shows that the true difference vector, $\mathbf{\delta}$, can be represented not only by the difference between vectors $\mathbf{t}_2$ and $\mathbf{t}_1$, but also by the sum of the vectors $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$. The latter system is advantageous because $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$ are always orthogonal whereas there may be a substantial correlation between $\mathbf{t}_1$ and $\mathbf{t}_2$.

These diagrams also provide a graphic interpretation of the regression coefficient a. Given (10), the standard deviation of the true dependent change scores may be written

$$\sigma_{\zeta_x} = (a_x - 1)\sigma_{t_{x1}} . \tag{14}$$

Again dropping the x's, we have

$$\sigma_{\zeta} = a\sigma_{t_1} - \sigma_{t_1} , \tag{15}$$

and

$$a = \frac{\sigma_{t_1} + \sigma_\zeta}{\sigma_{t_1}}, \tag{16}$$

where the sum in the right-hand term is understood as a vector addition. Equation (16) can be rewritten in terms of the cosine of the angle between $\mathbf{t}_1$ and $\mathbf{t}_2$,

$$a = (\sigma_{t_2}/\sigma_{t_1}) \cos \theta. \tag{17}$$

It is interesting to examine vector diagrams corresponding to various values of a. In Fig. 1A, the regression coefficient a is greater than $+1.00$. Here the projection of $\mathbf{t}_2$ on $\mathbf{t}_1$ is longer than $\mathbf{t}_1$ and the direction of the true difference vector, $\boldsymbol{\delta}$, shows that the true difference scores will be positively correlated with true scores on the first test.

In Fig. 1B, the coefficient a is unity. The projection of $\mathbf{t}_2$ on $\mathbf{t}_1$ is now exactly equal to $\mathbf{t}_1$; the true difference vector, $\boldsymbol{\delta}$, is collinear with the true independent gain score, $\boldsymbol{\gamma}$, and the true difference scores correlate zero with scores on the first test. Because the standard deviation of the true dependent change scores, σ_ζ, is now zero, the vector $\boldsymbol{\zeta}$ will have zero length.

Fig. 1C illustrates the condition where $a < 1$. The projection of $\mathbf{t}_2$ on $\mathbf{t}_1$ is shorter than $\mathbf{t}_1$; $\boldsymbol{\zeta}$ and $\mathbf{t}_1$ are oriented in opposite directions, and the direction of the true difference vector, $\boldsymbol{\delta}$, is such that true difference scores will have a negative correlation with true scores on the first test.

Estimating Components of True Difference Score

Procedures to estimate scores for the true independent change component, γ_{xi}, will be presented first. We require an estimate of the coefficient a, which was defined earlier as the regression of the true scores of the second testing on the true scores for the first testing. Dropping the x subscript,

$$a = \rho_{t_1 t_2}(\sigma_{t_2}/\sigma_{t_1}). \tag{18}$$

In using this expression we shall wish to exclude such trivial cases as $\rho_{t_2 t_1} = 0$ or $\sigma_{t_2} = 0$, for then $a = 0$ and the change score, γ_i, is equal to the true score on the second test, t_{i2}. We also wish to exclude the case where $\sigma_{t_1} = 0$, that is, where there is no true score variance on the first testing because the pretest is completely unreliable.

Into (18) we may substitute the familiar relationships between true and observed scores which may be found, for example, in Gulliksen [3], to yield

$$a = \frac{\rho_{x_1 x_2} \sigma_{x_2} \sqrt{\rho_{x_2 x_2}}}{\sqrt{\rho_{x_1 x_1} \rho_{x_2 x_2}}\, \sigma_{x_1} \sqrt{\rho_{x_1 x_1}}}, \tag{19}$$

where $\rho_{x_1 x_1}$ and $\rho_{x_2 x_2}$ are the population reliabilities of the first and second administrations of the test. On cancelling terms in (19),

$$a = \frac{\rho_{x_1x_2}\sigma_{x_2}}{\rho_{x_1x_1}\sigma_{x_1}}. \tag{20}$$

In empirical work with large samples of subjects we may find $\hat{a}$, which is an estimate of a, by substituting the appropriate sample statistics into (20).

$$\hat{a} = \frac{r_{x_1x_2}S_{x_2}}{r_{x_1x_1}S_{x_1}}. \tag{21}$$

Note that this is the ordinary regression of the observed scores of the second test on the observed scores of the first test divided by the reliability of the first test.

The next step in estimating independent gains scores for specific individuals requires substituting (1) and (2) into (8). As a result we have

$$\gamma_i = x_{i2} - ax_{i1} - (e_{i2} - ae_{i1}). \tag{22}$$

Let g_i be defined as an observed independent gains score in deviation form, i.e.,

$$g_i = x_{i2} - ax_{i1} . \tag{23}$$

By combining (22) and (23),

$$g_i = \gamma_i + (e_{i2} - ae_{i1}). \tag{24}$$

This equation resembles (6) which was derived for difference scores, for in both cases the measure of discrepancy or change (d_i and g_i) is seen to consist of a true score and an error term.

These considerations depend upon a knowledge of a. In empirical work we seldom know this value exactly; at best we have a more or less satisfactory estimate of a, defined as $\hat{a}$ in (21). If this estimate is satisfactory, it can be used to compute estimated independent gains scores, denoted $\hat{g}_i$. In deviation score form,

$$\hat{g}_i = x_{i2} - \hat{a}x_{i1} . \tag{25}$$

In raw score form,

$$\hat{G}_i = X_{i2} - \hat{a}X_{i1} . \tag{26}$$

When a base-free measure of change or discrepancy is required for correlational work, (26) or (25) may be used in conjunction with (21) to provide a satisfactory solution as long as the sample size is reasonably large.

The second problem is estimating scores for the true dependent change component, ζ_{xi} . Substituting (1) into (10) (and again dropping the x subscript),

$$\zeta_i = (a - 1)x_{i1} - (a - 1)e_{i1} . \tag{27}$$

Let w_i be defined as an observed dependent gains score in deviation form, i.e.,

$$w_i = (a - 1)x_{i1} . \tag{28}$$

By combining (27) and (28),

$$w_i = \zeta_i + (a - 1)e_{i1} , \tag{29}$$

which consists of a true score and an error term like (6) and (24).

If $\hat{a}$ is used as an estimate of a in (28), one may compute estimated dependent gains scores, denoted $\hat{w}_i$.

$$\hat{w}_i = (\hat{a} - 1)x_{i1} . \tag{30}$$

In raw score form,

$$\hat{W}_i = (\hat{a} - 1)X_{i1} . \tag{31}$$

In correlational studies of change, however, these estimated scores need not be computed, for they are a linear function of scores on the first test. The absolute value of the correlation of any variable with $\hat{w}_i$ or $\hat{W}_i$ will be exactly the same as the absolute value of the correlation of the same variable with the scores on the first test, that is, with x_{i1} or X_{i1} . The *sign* of the correlation may be different, however, for when any variable is correlated with $\hat{w}_i$ or $\hat{W}_i$ the sign of the correlation depends upon the value of $\hat{a}$. If $\hat{a} > 1$ the correlation will have the *same* sign and value as the correlation of this variable with scores on the first test. If $\hat{a} < 1$ the correlation will have the *opposite* sign (but the same absolute value) as the correlation of this variable with scores on the first test. If $\hat{a} = 1$ the correlation will be indeterminant because the true dependent change scores, ζ_i, will have zero variance.

Formulas for estimating the reliability of the true independent change score and the true dependent change score are given in a technical appendix. This appendix also contains formulas for computing the correlations between these two components of change and various external or predictor variables.

Independent Change Scores and Residual Scores

DuBois [2], Lacey [6], and Manning and DuBois [14] have also sought base-free measures of change. Our equations differ from theirs in utilizing true scores rather than observed scores in the measurement model. Some further comparison of their equations with ours is, perhaps, in order.

Earlier we noted in (20) that

$$a = \frac{\rho_{x_1x_2}\sigma_{x_2}}{\rho_{x_1x_1}\sigma_{x_1}}$$

and we may rewrite this as

$$a = \frac{1}{\rho_{x_1x_1}} b, \tag{32}$$

where b is the ordinary regression coefficient. Lacey [6] has constructed an autonomic lability score by using b in place of a. His score differs from ours in several other ways as well, but these differences should not affect correlations. Thus, Lacey uses standardized initial and final scores, divides his regressed scores by the standard error of estimate, and converts the result to a McCall's T score. Where z_{xi1} and z_{xi2} are standard scores on test X for subject i on occasions 1 and 2, Lacey's equation is

$$\text{ALS} = 50 + 10\left[\frac{z_{xi2} - r_{x_1x_2}z_{xi1}}{\sqrt{1 - r_{x_1x_2}^2}}\right]. \tag{33}$$

Since 50, 10, and $\sqrt{1 - r_{x_1x_2}^2}$ are all constants, (33) should give the same correlations as the following formula advocated by DuBois [2] and Manning and DuBois [14]:

$$z_{xi2\cdot1} = z_{xi2} - r_{x_1x_2}z_{xi1} . \tag{34}$$

Upon substituting deviation scores for the standard scores of (34),

$$x_{i2\cdot1} = x_{i2} - bx_{i1} , \tag{35}$$

where b is the ordinary regression coefficient for x_2 on x_1 . Equation (35) is the analogue of our (25): it differs only in using b instead of $\hat{a}$ as a regression coefficient. When the first test, x_1 , is perfectly reliable, there will be no difference in the correlations computed with (25) and (35), but as the initial position measurement becomes more unreliable the difference in the correlations computed from the two models begins to widen. Equation (35) may appear to give higher correlations in some cases of this type, but this should not be construed as an argument in its favor, for these higher correlations may be the result of ignoring the full effects of initial score unreliability on the slope of the regression line and removing too little of the variance of the initial position on the scores from the final testing. Equation (35) is less "base-free" than it ought to be, but (25) is more nearly optimal in this regard.

Discussion

These equations are offered as part of the continuing discussion of the fundamental logic of the measurement of change; a partial list of recent contributions to this field would include DuBois [2], Harris [4], Lacey [6], Lacey and Lacey [7], Lord [8, 9, 11, 12], Manning and DuBois [14], and McNemar [15]. Harris [5] has recently edited a volume of conference papers that adds to the growing literature on this topic.

Our methods of correcting for the unreliability of a control or base-line variable have antecedents and analogies in previous statistical articles. Berkson [1], for example, pointed out that errors of measurement in the control variable tend to flatten the slope of the regression line. Madansky

[13] discussed the fitting of regression lines under several different assumptions about errors of measurement. Stouffer [17, 18] provided methods for correcting unreliable control variables when computing partial correlations, and more recently Lord [10] devised a method for using the analysis of covariance with fallible control variables. Methods of this sort extend the usefulness of many classical statistical paradigms, but they have received, perhaps, insufficient attention in introductory books on psychological statistics.

As shown in Fig. 1 there are two ways of representing change or discrepancy: we may estimate the true difference score, δ, or we may estimate the true independent and dependent change scores, γ and ζ. Which method is preferable? The choice may depend largely on information already in the experimenter's hands and upon his research goals.

If the experimenter wishes the best possible estimate of the gain or loss experienced by specific individuals or groups, he may be advised to estimate the true difference scores (cf., Lord [8, 9, 11, 12]; McNemar [15]). Experimenters may also wish to use difference scores in correlational work when the knowledge of just these scores maximizes some predictive payoff. Again, difference scores may be the more useful when we know little about the factorial composition of the tests being studied. One usually prefers to regress the factorially simpler test out of the factorially more complex one; but if there is some uncertainty about which is which, it may be better to use a difference score, for the latter is symmetrical. Except for the change in sign, it does not matter whether t_1 is subtracted from t_2 or t_2 from t_1, for the reliabilities will be the same and the correlation with other measures will have the same absolute value. It may matter very much whether t_1 is regressed out of t_2 or t_2 out of t_1 for the reliabilities may differ and so may the correlation of the change scores with other variables.

Finally, observed difference scores may be preferred in some sorts of work because they are properties of the subject himself whereas the estimated scores, particularly the independent and dependent change scores, reflect not only the properties of the person, but the properties of the group used to make the estimate.* Whenever an experimental or administrative program requires that change scores characterize the person while he is transferred from one group to another, the difference score may be the more useful statistic.

There are many research problems, however, in which the experimenter computes correlation coefficients in order to refine his analysis of the determinants of a discrepancy or of a change produced by some treatment. In

*It is possible to compute estimated difference scores or estimated independent change scores separately for each of N individuals by subjecting each person to the experimental conditions n times. In this case the a of (18), (19), and (20) and the $\hat{a}$ of (21), (25), and (26) is calculated separately for each person on the n replications of the experiment. The relationship of this sort of measure to the group statistics version of a and $\hat{a}$ is an interesting theoretical problem but beyond the scope of the present paper (cf. Lacey, [6]).

problems of this sort it is often important to learn which variables determine the subject's position on the initial base-line or control score and which variables contribute to change in ways that are independent of this initial position. Experimenters often explore many independent variables in an effort to predict change. When n such variables are available, the experimenter may wish to construct a $2 \times n$ table showing the correlation of each predictor with estimates of the independent and dependent true change scores. He may also wish to correct these correlations for attenuation in the ordinary way by using the reliabilities of the predictor variables and the reliabilities of the components of change as these appear in a technical appendix to the present paper.

Certain problems arise when one of the components of the test used to measure change is also used as a predictor variable. If, for example, the base-line measure x_1 is correlated with the true independent change score $\hat{g}_x$, and if the measures are less than perfectly reliable—as is usually the case—there may appear to be a negative relationship between the two variables. Nevertheless, the correlation between the *true* scores for the two variables was assumed to be zero in deriving $\hat{g}_x$. Therefore, any relationship between the observed scores stems entirely from correlated errors of measurement and should be disregarded. The correlation between the base-line measure x and the estimated true dependent change score $\hat{w}_x$ will be $+1.00$ when $\hat{a} > 1$, and -1.00 when $\hat{a} < 1$. It will be indeterminant when $\hat{a} = 1$, for at this value $\hat{w}_x$ will have no true score variance. The latter relationships do not involve correlated error and are thus independent of the (identical) reliabilities of x_1 and $\hat{w}_x$ (assuming that these reliabilities differ significantly from zero).

The correlation between the base-line measure and the observed difference score, d_x, is affected both by correlated errors of measurement and by real relationships between the true score components, t_1 and δ. Typically one wishes to obtain a correlation that is free from the effects of correlated error, and there are two ways of doing this. One may, for example, administer two parallel forms of test X to obtain the first or base-line measurement. The errors of measurement of parallel forms are independent (cf. Gulliksen [3]), and therefore one of the two forms may be used to calculate the difference score, while the other is used to obtain the correlation of the difference score with initial position. The alternative method involves correcting the correlation between the difference score and the base-line measurement for correlated errors. Thomson [19, 20] devised an equation for this purpose and Zieve [21] provided an alternative form, which is given in the technical appendix as equation (62). The Thomson and Zieve equations, however, correct not only for correlated error but for attenuation as well, and this second correction should be taken into account when comparing the initial position with various independent variables for their relative merits in predicting a difference score. One should either correct all correlations for attenuation before making such comparisons or modify the Thomson and Zieve equations by multiplying them by the geometric mean of the reliabilities of the two variables involved. A corrected version of the Zieve equation appears in the technical appendix as equation (63).

One may conclude that once a treatment is understood well enough to know whether it makes subjects simpler or more complex and once a study has been designed so that subjects in the treatment group can be compared meaningfully for response to this treatment, then the two new measures developed in this paper are likely to be useful and appropriate statistics for the study of change.

Technical Appendix

This appendix will offer reliabilities for estimated independent and dependent gains scores and formulas for the correlation of scores of this type with other variables. These expressions will be compared with analogous formulas for residual scores and difference scores in order to clarify various properties of independent gain.

We may develop a reliability for the estimated independent gains scores, $\hat{g}$ in (25) or $\hat{G}$ in (26), in terms of population values and then use sample statistics to estimate that reliability in specific empirical problems. Conventionally, we define the reliability as the proportion of true score variance in the total variance of the gain scores,

$$\rho_{gg} = \frac{\sigma_\gamma^2}{\sigma_g^2}, \tag{36}$$

where γ is defined in (8) and g in (23). The variance of the true scores may be found with the aid of (8) to be

$$\sigma_\gamma^2 = E[(t_2 - at_1)^2]. \tag{37}$$

Expanding the right-hand side of (37) and substituting observed for true scores,

$$\sigma_\gamma^2 = \rho_{x_2x_2}\sigma_{x_2}^2 - 2a\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + a^2\rho_{x_1x_1}\sigma_{x_1}^2 . \tag{38}$$

If we substitute for a the value given in (20),

$$\sigma_\gamma^2 = \frac{\sigma_{x_2}^2(\rho_{x_1x_1}\rho_{x_2x_2} - \rho_{x_1x_2}^2)}{\rho_{x_1x_1}}. \tag{39}$$

The variance of the estimated scores may be obtained in a like manner from (23).

$$\sigma_g^2 = E[(x_2 - ax_1)^2]. \tag{40}$$

Expanded, this becomes

$$\sigma_g^2 = \sigma_{x_2}^2 - 2a\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + a^2\sigma_{x_1}^2 \, . \tag{41}$$

Substituting for a, as before,

$$\sigma_g^2 = \frac{\sigma_{x_2}^2(\rho_{x_1x_1}^2 - 2\rho_{x_1x_2}^2\rho_{x_1x_1} + \rho_{x_1x_2}^2)}{\rho_{x_1x_1}^2}. \tag{42}$$

Combining (36), (39), and (42), we obtain an expression for the reliability in population notation

$$\rho_{gg} = \frac{\rho_{x_1x_1}(\rho_{x_1x_1}\rho_{x_2x_2} - \rho_{x_1x_2}^2)}{\rho_{x_1x_1}^2 - 2\rho_{x_1x_2}^2\rho_{x_1x_1} + \rho_{x_1x_2}^2}. \tag{43}$$

When the sample size is reasonably large, the sample correlations and reliabilities may be substituted into (43) to obtain an estimate of the reliabilities of the gains scores.

$$\hat{\rho}_{gg} = r_{gg} = \frac{r_{x_1x_1}(r_{x_1x_1}r_{x_2x_2} - r_{x_1x_2}^2)}{r_{x_1x_1}^2 - 2r_{x_1x_2}^2 r_{x_1x_1} + r_{x_1x_2}^2}. \tag{44}$$

The adequacy of (44) with small samples has not yet been investigated.

Inspection of (43) and (44) reveals some important properties of independent change scores. When, for example, the reliabilities of both component variables are unity, the independent change scores will also have unit reliability no matter what the correlation between the first and second testing may be. When the product of the reliabilities of the two components equals the squared correlation between them, the independent change scores will have zero reliability. When there is no correlation between the two components, the reliability of the change is the same as the reliability of the test on its second administration.

The reliability of the estimated dependent gains score, $\hat{w}$ in (30) or $\hat{W}$ in (31), is the same as the reliability of the first test or basal measure. Consider that

$$\rho_{ww} = \frac{\sigma_\zeta^2}{\sigma_w^2}. \tag{45}$$

By squaring both sides of (14) (and dropping the x subscript)

$$\sigma_\zeta^2 = (a - 1)^2\sigma_{t_1}^2 \, . \tag{46}$$

The variance of the observed scores may be found with the aid of (28). It is

$$\sigma_w^2 = E[(a - 1)^2 x_1^2], \tag{47}$$

which becomes

$$\sigma_w^2 = (a - 1)^2\sigma_{x_1}^2 \, . \tag{48}$$

Therefore

$$\rho_{ww} = \frac{\sigma_\zeta^2}{\sigma_w^2} = \frac{(a-1)^2\sigma_{t_1}^2}{(a-1)^2\sigma_{x_1}^2} = \frac{\sigma_{t_1}^2}{\sigma_{x_1}^2} = \rho_{x_1x_1} . \tag{49}$$

Comparisons with Other Reliabilities

Alternative versions of (43) and (44) might have been derived from (36), (38), and (41) in place of (36), (39), and (42). The new version of (43) would be

$$\rho_{gg} = \frac{\rho_{x_2x_2}\sigma_{x_2}^2 - 2a\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + a^2\rho_{x_1x_1}\sigma_{x_1}^2}{\sigma_{x_2}^2 - 2a\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + a^2\sigma_{x_1}^2}. \tag{50}$$

In terms of sample statistics,

$$r_{gg} = \frac{r_{x_2x_2}S_{x_2}^2 - 2\hat{a}r_{x_1x_2}S_{x_1}S_{x_2} + \hat{a}^2r_{x_1x_1}S_{x_1}^2}{S_{x_2}^2 - 2\hat{a}r_{x_1x_2}S_{x_1}S_{x_2} + \hat{a}^2S_{x_1}^2}. \tag{51}$$

Equation (51) is equivalent to (44) but it is more instructive when compared to the usual formula for the reliability of observed *difference* scores, which appears in Lord [12], Manning and DuBois [14], and McNemar [15]. This equation—in terms of sample statistics—is

$$r_{dd} = \frac{r_{x_2x_2}S_{x_2}^2 - 2r_{x_1x_2}S_{x_1}S_{x_2} + r_{x_1x_1}S_{x_1}^2}{S_{x_2}^2 - 2r_{x_1x_2}S_{x_1}S_{x_2} + S_{x_1}^2}. \tag{52}$$

Equations (51) and (52) have much the same form except for the appearance in (51) of an $\hat{a}$ or an $\hat{a}^2$ in every term containing S_{x_1} or $S_{x_1}^2$. Clearly then, when $\hat{a} = 1$, the independent gains score is the same as the difference score and $r_{gg} = r_{dd}$.

The reliability for the regressed score used by Lacey and DuBois has been given by Manning and DuBois [14] and by Messick and Hills [16]. In sample statistics this is

$$r_{x_2\cdot 1x_2\cdot 1} = \frac{r_{x_2x_2} - 2r_{x_1x_2}^2 + r_{x_1x_2}^2r_{x_1x_1}}{1 - r_{x_1x_2}^2}. \tag{53}$$

It is instructive to compare (53) with the reliability of our independent gains score, given in (44). If we assumed, for example, that the "pretest" were perfectly reliable we could set $r_{x_1}r_{x_1}$ at unity, then (44) and (53) would reduce to the same form, namely

$$r_{x_2\cdot 1x_2\cdot 1} = r_{gg} = \frac{r_{x_2x_2} - r_{x_1x_2}^2}{1 - r_{x_1x_2}^2}. \tag{54}$$

This result also follows from (32).

The two reliabilities r_{gg} and $r_{x_2\cdot 1x_2\cdot 1}$ begin to diverge as the reliabilities of the initial scores diminish. At first the differences are not large. When, for example, $r_{x_1x_1} = .900$, $r_{x_2x_2} = .880$, and $r_{x_1x_2} = .389$, then the residual score reliability $r_{x_2.1x_2.1} = .840$, and the estimated independent gains score reliability $r_{gg} = .740$. But when, as in the study reported by Manning and DuBois ([14], p. 303), $r_{x_1x_1} = .490$, the two reliabilities diverge more sharply. If $r_{x_2x_2} = .880$ and $r_{x_1x_2} = .389$ again, we find $r_{x_2.1x_2.1} = .770$ while $r_{gg} = .564$. The regressed score is not to be preferred on account of its higher apparent reliability, for this high reliability was obtained by regressing out too little of the variance of the initial testing from scores on the final testing.

Correlational Formulas

An expression for the correlation of two independent change scores, can be derived by utilizing equations like (42) for the variances and by expanding the following expression for the covariance:

$$\text{cov}_{g_xg_y} = E[(x_2 - a_xx_1)(y_2 - a_yy_1)]. \tag{55}$$

The actual equation, in population notation, is

$$\rho_{g_xg_y} = \frac{\rho_{x_2y_2}\rho_{x_1x_1}\rho_{y_1y_1} - \rho_{x_1x_2}\rho_{y_1y_1}\rho_{x_1y_2} - \rho_{y_1y_2}\rho_{x_2y_1}\rho_{x_1x_1} + \rho_{x_1x_2}\rho_{y_1y_2}\rho_{x_1}}{\sqrt{\rho_{x_1x_1}^2 - 2\rho_{x_1x_2}^2\rho_{x_1x_1} + \rho_{x_1x_2}^2}\sqrt{\rho_{y_1y_1}^2 - 2\rho_{y_1y_2}\rho_{y_1y_1} + \rho_{y_1y_2}^2}}. \tag{56}$$

The correlation between an independent change score and an ordinary "stationary" test score, k, has a simpler form, viz.,

$$\rho_{g_xk} = \frac{\rho_{x_2k}\rho_{x_1x_1} - \rho_{x_1k}\rho_{x_1x_2}}{\sqrt{\rho_{x_1x_1}^2 - 2\rho_{x_1x_2}^2\rho_{x_1x_1} + \rho_{x_1x_2}^2}}. \tag{57}$$

Equations (56) and (57) assume that tests X and Y have significant true score variance on their first administration and that the error of measurement associated with the use of $\hat{g}_x$ in place of γ_x is uncorrelated with the errors of measurement in $\hat{g}_y$ and k. This assumption is violated, of course, in the special case where an estimated independent gains score is correlated with either of its component parts.

If we set the reliability of the first of the two test administrations equal to unity in (56) and (57), we will obtain formulas for the correlation of two residual scores with each other and the formulas for the correlation of a residual score with an external variable.

$$\rho_{x_2.1y_2.1} = \frac{\rho_{x_2y_2} - \rho_{x_1x_2}\rho_{x_1y_2} - \rho_{y_1y_2}\rho_{x_2y_1} + \rho_{x_1x_2}\rho_{y_1y_2}\rho_{x_1y_1}}{\sqrt{1 - \rho_{x_1x_2}^2}\sqrt{1 - \rho_{y_1y_2}^2}}. \tag{58}$$

$$\rho_{x_2.1k} = \frac{\rho_{x_2k} - \rho_{x_1k}\rho_{x_1x_2}}{\sqrt{1 - \rho_{x_1x_2}^2}}. \tag{59}$$

Equation (59) is the "part correlation" used by DuBois [2] in his study of change. Equations (58) and (59) may yield higher correlations than (56) and (57) when applied to the same data, but this is not—as we have already suggested—an argument in their favor.

A formula for the correlation of two *difference* scores, d_x and d_y can easily be obtained. It is

$$\rho_{d_x d_y} = \frac{\rho_{x_2 y_2}\sigma_{x_2}\sigma_{y_2} - \rho_{x_1 y_2}\sigma_{x_1}\sigma_{y_2} - \rho_{x_2 y_1}\sigma_{x_2}\sigma_{y_1} + \rho_{x_1 y_1}\sigma_{x_1}\sigma_{y_1}}{\sqrt{\sigma_{x_2}^2 - 2\rho_{x_1 x_2}\sigma_{x_1}\sigma_{x_2} + \sigma_{x_1}^2}\,\sqrt{\sigma_{y_2}^2 - 2\rho_{y_1 y_2}\sigma_{y_1}\sigma_{y_2} + \sigma_{y_2}^2}}. \tag{60}$$

The corresponding formula for correlating an observed difference score with an ordinary variable, k, is

$$\rho_{d_x k} = \frac{\rho_{x_2 k}\sigma_{x_2} - \rho_{x_1 k}\sigma_{x_1}}{\sqrt{\sigma_{x_2}^2 - 2\rho_{x_1 x_2}\sigma_{x_1}\sigma_{x_2} + \sigma_{x_1}^2}}. \tag{61}$$

Again we call attention to the assumption that the error components in the variables being correlated are independent. This is patently false when difference scores are correlated with initial and final scores on the same test. The correlation between the true difference score and true initial position will not usually be zero, however, but correlated errors will still affect our estimate of this value from raw scores. Thomson [19, 20] developed an equation for finding the correlation of true scores, and Zieve [21] provided an alternative form which is

$$\rho_{t_1 \delta} = \frac{\rho_{x_1 d} + (\sigma_{x_1}/\sigma_d)(1 - \rho_{x_1 x_1})}{\sqrt{\rho_{x_1 x_1}}\,\sqrt{\rho_{dd}}}, \tag{62}$$

where t_1 is the true score for the initial position and δ is the true difference score. This equation, like those proposed by Thomson, is a correlation between true scores and is thus corrected, not only for correlated error, but for attenuation as well. If correction for attenuation is not desired, $\rho_{t_1 \delta}$ may be multiplied by the geometric mean of the reliabilities of the two variables involved. Equation (62) now becomes

$$_c\rho_{x_1 d} = \rho_{x_1 d} + (\sigma_{x_1}/\sigma_d)(1 - \rho_{x_1 x_1}), \tag{63}$$

where x_1 is the observed initial score and d is the observed difference score and $_c\rho_{x_1 d}$ refers to a correlation between these variables that has been corrected for correlated error.

An Interpretation of the "Law of Initial Values"

The correlation between the initial position and the difference score is related to a, the regression coefficient for independent change scores. Using (61) we have:

$$\rho_{x_1 d} = \frac{(a - 1)\sigma_{x_1}\rho_{x_1 x_1}}{\sqrt{\sigma_{x_2}^2 - 2\rho_{x_1 x_2}\sigma_{x_1}\sigma_{x_2} + \sigma_{x_1}^2}}. \tag{64}$$

Clearly, when a is greater than one there will be a positive correlation between initial position and change. When a is equal to one this correlation will be zero, and when it is less than one the correlation will be negative.

In explicating these relationships it may be helpful to consider a as the product of two ratios

$$(65) \qquad a = \left(\frac{\rho_{x_1x_2}}{\rho_{x_1x_1}}\right)\left(\frac{\sigma_{x_2}}{\sigma_{x_1}}\right).$$

If the system under consideration changes cumulatively, then the greater the time lapse between the first and second testings the lower the first of these ratios will be. This occurs because the reliability of the first testing remains constant while the correlation between the first and second testing decreases with the passage of time. If under these circumstances the population grows more heterogeneous with time, the increase in σ_{x_2} may make up for the decrease in $\rho_{x_1x_2}$ in such a way that $a_x = 1$. In this case there will be no correlation between change and initial position. But some psychological and biological systems show a conservative tendency in that the standard deviation of a trait tends to remain constant even while real changes are taking place (cf., Lord [12]). As a result

$$\frac{\rho_{x_1x_2}}{\rho_{x_1x_1}} < \frac{\sigma_{x_2}}{\sigma_{x_1}} \cong 1,$$

and $a < 1$. In a system of this sort the correlation between initial position and the difference score will tend to be negative, and a "law of initial values" as discussed by Lacey and Lacey [7] will appear.

REFERENCES

[1] Berkson, J. Are there two regressions? *J. Amer. statist. Ass.*, 1950, 45, 164–180.

[2] DuBois, P. H. *Multivariate correlational analysis*. New York: Harper, 1957.

[3] Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

[4] Harris, C. W. Some problems in the description of change. *Educ. psychol. Measmt*, 1962, 22, 303–319.

[5] Harris, C. W. (Ed.) *Problems in measuring change*. Madison: Univ. Wisconsin Press, 1963.

[6] Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. New York Acad. Sci.*, 1956, 67, 123–164.

[7] Lacey, J. I. and Lacey, B. C. The law of initial value in the longitudinal study of autonomic constitution: Reproducibility of autonomic responses and response patterns over a four year interval. In W. M. Wolf (Ed.), Rhythmic functions in the living system. *Ann. New York Acad. Sci.*, 1962, 98, 1257–1290.

[8] Lord, F. M. The measurement of growth. *Educ. psychol. Measmt*, 1956, 16, 421–437.

[9] Lord, F. M. Further problems in the measurement of growth. *Educ. psychol. Measmt*, 1958, 18, 437–451.

[10] Lord, F. M. Large-sample covariance analysis when the control variable is fallible. *J. Amer. statist. Ass.*, 1960, 55, 307–321.

[11] Lord, F. M. Measurement of academic growth—statistical problems. Res. Memo. 62–3, Princeton, N. J.: Educ. Test. Serv., 1962.

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

[13] Madansky, A. The fitting of straight lines when both variables are subject to error. *J. Amer. statist. Ass.*, 1959, **54**, 173–205.
[14] Manning, W. H. and DuBois, P. H. Correlational methods in research on human learning. *Percept. mot. Skills*, 1962, **15**, 287–321.
[15] McNemar, Q. On growth measurement. *Educ. psychol. Measmt*, 1958, **18**, 47–55.
[16] Messick, S. and Hills, J. K. Objective measurement of personality: cautiousness and intolerance of ambiguity. *Educ. psychol. Measmt*, 1960, **20**, 685–698.
[17] Stouffer, S. A. Evaluating the effect of inadequately measured variables in partial correlation analysis. *J. Amer. statist. Ass.*, 1936, **31**, 348–360.
[18] Stouffer, S. A. Reliability coefficients in a correlation matrix. *Psychometrika*, 1936, **1**, 17–20.
[19] Thomson, G. H. A formula to correct for the effect of errors of measurement on the correlation of initial values with gains. *J. exp. Psychol.*, 1924, **7**, 321–324.
[20] Thomson, G. H. An alternative formula for the true correlation of initial values with gains. *J. exp. Psychol.*, 1925, **8**, 323–324.
[21] Zieve, L. Note on the correlation of initial scores with gains. *J. educ. Psychol.*, 1940, **31**, 391–394.

Manuscript received 7/16/65
Revised manuscript received 1/27/66