Journal of Management
1994, Vol. 20, No. 3, 683-689

# Regression Analysis as an Alternative to Difference Scores

Jeffrey R. Edwards
*University of Michigan*

For nearly 40 years, it has been asserted (see, e.g. Cronbach, 1958, 1992; Cronbach & Furby, 1970; Cronbach & Gleser, 1953; Edwards, 1994; Edwards & Cooper, 1990; Johns, 1981; Wall & Payne, 1973; Werts & Linn, 1970) that difference scores suffer from various methodological problems (about which more anon). In their position statement, Tisak and Smith argue that some of these problems have been overstated and suggest alternative procedures (e.g., the expanded difference equation) intended to overcome certain problems while maintaining the use of difference scores. Although certain points made by Tisak and Smith have merit, they minimize or overlook several important problems with difference scores, and their recommended procedures fail to overcome these problems. I will elaborate my position according to the two primary issues addressed by Tisak and Smith, the reliability and validity of difference scores. I will then note shortcomings with the Tisak and Smith procedure and contend that the regression procedure described by Edwards (1994) mitigates or avoids arguable problems with difference scores, but permits comprehensive tests of conceptual models that difference scores are intended to represent.

## Reliability

In defense of the reliability of difference scores, Tisak and Smith argue that difference scores are not inherently unreliable, but may prove unreliable when the component measures comprising the difference are unreliable and positively correlated. Tisak and Smith also point out that, unlike bivariate difference scores, profile similarity indices are often based on composite (multi-item) multiple source measures. Because of this, profile similarity indices are likely to yield higher reliability estimates than bivariate difference scores.

As pointed out by Tisak and Smith, it is undeniable that the reliability of any measure is ultimately an empirical question that should be addressed on a study-by-study basis. However, the primary message of Johns (1981) and others is that the conditions under which difference scores are unreliable (i.e., positively correlated component measures with modest reliabilities) are common in empirical research. This is not surprising, given that difference score

Direct all correspondence to: Jeffrey R. Edwards, University of Michigan, School of Business Administration, Ann Arbor, MI 4809-1234.

components are usually measured with the same instrument and often represent constructs that should be positively correlated on conceptual grounds. For example, Schneider (1987) argues that people gravitate toward work settings that are similar to themselves, thereby generating a positive correlation between measures of the person and job. Because of this, it is reasonable to assert *a priori* that difference scores may well exhibit poor reliabilities. Furthermore, the reliability of a difference score should be evaluated not only in an absolute sense, but also in relation to viable alternatives, such as using both component measures jointly in multiple regression analysis (Edwards, 1994; Edwards & Cooper, 1990). If a difference score exhibits adequate reliability, then it is almost certain that its component measures will exhibit superior reliabilities, indicating that the latter should be used in place of the former (Edwards, 1994).

Unlike bivariate difference scores, profile similarity indices (e.g., $D^2$) will often exhibit reliabilities that are substantially larger than their component measures. This is due in part to the number of dimensions involved in the calculation of the index, which has a dramatic impact on its estimated reliability (Nunnally, 1978). For example, if 10 squared differences exhibiting reliabilities of .50 and intercorrelations of .10 were standardized and summed to form $D^2$, the reliability of the resulting index would be .74. Studies using profile similarity indices (e.g., Caldwell & O'Reilly, 1990; Chatman, 1991; Dougherty & Pritchard, 1985; Rounds, Dawis & Lofquist, 1987) often incorporate a much larger number of dimensions, virtually guaranteeing that the index will demonstrate high reliability.

Although profile similarity indices may yield high reliability estimates, the interpretation of these estimates can be problematic. Reliability is typically defined as the proportion of true score variance in a measure, or the squared correlation between a measure and its associated underlying construct (Lord & Novick, 1968; Nunnally, 1978). Unless the items comprising a measure share a common meaning, it is difficult to define the construct underlying the measure, and the interpretation of the reliability of the measure therefore becomes suspect (Gerbing & Anderson, 1988; Hattie, 1985; Wolins, 1982). In my experience, the items typically comprising profile similarity indices represent conceptually distinct dimensions and, hence, do not share a common meaning. For example, dimensions measured by Chatman (1991) included aggressiveness, risk taking, precision, and social responsibility, those measured by Dougherty and Pritchard (1985) included making presentations, keeping records, and providing written advice to clients, and those measured by Smith and Tisak (1993) included data entry, obtaining information from clients, and interpreting company policies and procedures. In these cases, it seems difficult to define a construct that encompasses such diverse dimensions. Although it may be argued that indices that combine diverse dimensions represent similarity in a global sense, Cronbach and Gleser (1953) and Lykken (1956) have forcefully argued that similarity is meaningful only in terms of specific dimensions, not as a general quality. Without a clear definition of the construct underlying a profile similarity index, the concept of a "true score" is meaningless, and the reliability of the index becomes moot.

## Validity

Tisak and Smith acknowledge several problems pertaining to the validity of difference scores, such as ambiguous interpretation, confounding the effects of their component measures, and failure to explain variance beyond their component measures. Nonetheless, they assert that these problems do not provide sufficient justification to abandon difference scores *a priori,* arguing that the severity of each problem should be assessed empirically within the context of the data. Tisak and Smith further argue that, even when evidence for these problems is found (e.g., a difference score explains less variance than its component measures), the utility of difference scores remains a value judgment for the researcher.

Tisak and Smith are correct in pointing out that the severity of problems regarding the validity of difference scores can be assessed empirically. For example, the degree to which an algebraic difference explains less variance than its components can be assessed by comparing the $R^2$ from Equation 4 to that obtained from Equation 2, using a conventional F-test (Edwards, 1994). If Equation 4 explains significantly more variance than Equation 2, then the functional form associated with the algebraic difference (i.e., equal but opposite effects for the two component measures) is rejected, and the form indicated by Equation 4 should be preferred. If Equation 4 does not explain significantly more variance than Equation 2, then the functional form for the algebraic difference may be considered tenable (in both cases, it is also necessary to ensure that the overall $R^2$ is significant and no significant higher-order terms are found, thereby establishing that a linear equation adequately represents the functional form relating the component measures to the outcome; see Edwards, 1994). In neither case is it necessary or desirable to resort to Equation 2 once Equation 4 has been estimated. Moreover, the F-test comparing the $R^2$ values from Equations 2 and 4 can be replaced by a direct test of whether $\beta_1$ and $\beta_2$ in Equation 4 are equal in magnitude but opposite in sign (Cohen & Cohen, 1983, pp. 479-480), which makes Equation 2 superfluous (for analogous tests pertaining to absolute and squared difference scores, see Edwards, 1994).

The use of Equation 4 also avoids other problems regarding the validity of algebraic difference scores. For example, the interpretational ambiguity created by combining the component measures into a single composite is eliminated, given that the component measures are used as separate predictors. In addition, the effects of the component measures are no longer confounded, because separate coefficients are obtained for each measure. Of course, these advantages also pertain when Equation 6 is used in place of Equation 5, or when the piecewise linear equation described by Edwards (1994) is used instead of an absolute difference.

Tisak and Smith also attempt to bolster the validity of difference scores by arguing that they capture something distinct from their component measures. However, because difference scores are simply composites of their component measures, they cannot contain information beyond that available when these measures are considered *jointly* (Johns, 1981). Furthermore, as shown by

comparing regression equations using difference scores (e.g., Equations 2 and 5) to their unconstrained counterparts (Equations 4 and 6, respectively), the former equations are simply special cases of the latter. Because of this, it is logically impossible for equations using difference scores as predictors to capture anything beyond that represented by equations using difference score components. Moreover, equations using difference score components can capture theoretically meaningful effects that cannot be detected when equations relying on difference scores are used (for examples, see Edwards, 1994; Edwards & Harrison, 1993).

## The Viability of the Tisak and Smith Procedure

Tisak and Smith contend that tests comparing constrained regression equations using difference scores (e.g., Equation 5) to their unconstrained counterparts (e.g., Equation 6) are "inherently unfair," given that difference score equations contain only one parameter. As an alternative, they propose a generalized difference equation, Equation 7, that uses an algebraic and a squared difference as predictors.

There are two fundamental problems with the generalized difference equation proposed by Tisak and Smith. First, beyond the argument that it "maintains the idea of a difference between the components," there is no apparent conceptual justification for Equation 7. The central issue in testing the effects of congruence (i.e., fit, similarity, or agreement) is not whether a difference score is used in the equation, but whether the functional form relating the component measures to the outcome is consistent with that represented by the difference score. This cannot be determined by merely inserting a difference score into the equation, because a significant coefficient on a difference score can be generated by a substantial variety of functional forms, only one of which is consistent with that represented by the difference score itself (for examples of this, see Edwards, 1994; Edwards & Harrison, 1993). Further inspection of Equation 7 reveals that it is conceptually similar to Equation 5, but can depict minima at locations other than the point where X and Y are equal (specifically, if $\beta_1$ in Equation 7 is positive, the minimum is shifted to the region where $X < Y$, whereas if $\beta_1$ is negative, the minimum is shifted to the region where $X > Y$).

Second, when compared to Equation 5, Equation 7 simply replaces one set of constraints on Equation 6 with another (for the ensuing discussion, it is assumed that all coefficients in Equation 6 are estimated simultaneously). In particular, Equations 5 and 7 both impose the constraints $\beta_4 = \beta_5$ and $\beta_3 = -2\beta_4$. However, whereas Equation 5 constrains $\beta_1 = \beta_2 = 0$, Equation 7 constrains $\beta_1 = -\beta_2$. To test the constraints imposed by Equation 7, it is necessary to estimate Equation 6 and test the increment in $R^2$ yielded by Equation 6 over Equation 7 or, equivalently, directly test whether the coefficients from Equation 6 follow the pattern corresponding to Equation 7 (Dwyer, 1983). If the constraints imposed by Equation 7 are rejected and the set of cubic terms composed of $X_1$ and $Y_1$ is not significant (Edwards, 1994), then interpretation should focus on Equation 6, using procedures described by

Edwards and Parry (1993). If the constraints are not rejected, then the functional form corresponding to Equation 7 may be considered tenable. This, however, does not mean that Equation 7 should then be estimated, because the functional form relating the component measures to the outcome can be obtained directly from Equation 6. Furthermore, additional information that could be found by estimating Equation 7, such as its $R^2$ and coefficient estimates, can be calculated from the results of Equation 6, provided the constraints imposed on Equation 6 to yield Equation 7 are known (e.g., Johnston, 1984). The primary utility of Equation 7 is that it allows a researcher to construct hypotheses regarding the pattern of coefficients from Equation 6 that would yield support for the functional form corresponding to it. However, once Equation 6 has been estimated, it is unnecessary and redundant to then estimate Equation 7.

Tisak and Smith propose two generalizations of Equation 7, one using the sum of algebraic and squared differences across multiple dimensions (i.e., Equation 8), and another adding a second set of analogous summed difference measures (i.e., Equation 9). Unfortunately, Equations 8 and 9 simply compound the problems associated with Equation 7. This can be seen by considering the following equation, which is an expanded version of Equation 8:

$$\begin{aligned} Z = {} & \beta_0 + \beta_1(X_1 - Y_1) + \beta_2(X_1 - Y_1)^2 \\ & + \beta_1(X_2 - Y_2) + \beta_2(X_2 - X_2)^2 \\ & + \beta_1(X_2 - Y_3) + \beta_2(X_3 - Y_3)^2 + e \end{aligned} \tag{10}$$

As Equation 10 shows, Equation 8 imposes the same constraints as Equation 7 on the algebraic and squared differences corresponding to each dimension. Moreover, Equation 8 constrains coefficients across dimensions, such that the coefficients on each algebraic difference are the same, and the coefficients on each squared difference are the same. Conceptually, this implies that the functional form relating each paired $X_i$ and $Y_i$ to the outcome is the same, regardless of the substantive distinctions among the dimensions. Obviously, such an elaborate set of constraints should be tested empirically, not simply imposed on the data. This can be accomplished using the following equation, which is a generalization of Equation 6:

$$\begin{aligned} Z = {} & \beta_0 + \beta_1X_1 + \beta_2Y_1 + \beta_3X_1Y_1 + \beta_4X_1^2 + \beta_5Y_1^2 \\ & + \beta_6X_2 + \beta_7Y_2 + \beta_8X_2Y_2 + \beta_9X_2^2 + \beta_{10}Y_2^2 \\ & + \beta_{11}X_3 + \beta_{12}Y_3 + \beta_{13}X_3Y_3 + \beta_{14}X_3^2 + \beta_{15}Y_3^2 + e \end{aligned} \tag{11}$$

The constraints imposed by Equation 8 can be evaluated by testing the increment in $R^2$ yielded by Equation 11 or by directly testing whether the coefficients obtained from Equation 11 conform to the pattern associated with Equation 8. As before, once Equation 11 has been estimated, it is unnecessary to estimate Equation 8, regardless of whether the constraints imposed by Equation 8 are supported. An analogous unconstrained equation corresponding to Equation 9 can be derived and tested in a similar manner.

Estimating equations such as Equation 11 carries the obvious disadvantage of requiring large samples, particularly when the number of dimensions is large. However, the additional degrees of freedom provided by Equation 8 over 11 are obtained only by imposing constraints that are highly restrictive and, based on prior work with similar equations (Edwards, 1993), are unlikely to receive empirical support. Fortunately, this disadvantage is ameliorated when the dimensions are conceptually homogeneous, in which case the $X_i$ and $Y_i$ should be summed prior to analysis to form composite X and Y scales. For example, if the Role Conflict items described by Tisak and Smith represent a single underlying construct and satisfy the requirements for unidimensional measurement (Gerbing & Anderson, 1988; Hattie, 1985), then scales representing the employee's and supervisor's responses should be constructed by summing the corresponding items, and these scales should be used in Equation 6. When a larger number of dimensions is involved, as in studies using profile similarity indices (Caldwell & O'Reilly, 1990; Chatman, 1991; Dougherty & Pritchard, 1985; Rounds et al., 1987), it is likely that the dimensions can be distilled into a more parsimonious set (O'Reilly, Chatman & Caldwell, 1991) which, provided sample sizes were adequate, would permit the use of an equation such as Equation 11.

## Applications of the Edwards Procedure

The preceding discussion has contended that the aforementioned methodological problems with difference scores can be mitigated or avoided by applying the regression procedure described by Edwards (1994). The merits of this procedure over difference scores is not simply a matter of intellectual debate, but has also been demonstrated empirically. For example, Edwards (1994) found that, on average, when the constraints imposed by the algebraic, absolute, and squared differences between actual and desired job attributes were relaxed, the variance explained in job satisfaction nearly tripled. Similarly, Edwards and Harrison (1993) reanalyzed data from the classic P-E fit study conducted by French, Caplan and Harrison (1982) and found that, when the constraints imposed by the difference scores used by French et al. (1982) were relaxed, the variance explained in strain more than doubled. In both studies, the unconstrained regression equations indicated three-dimensional surfaces that were theoretically meaningful but notably more complex than the simplistic two-dimensional functions corresponding to bivariate difference scores. Furthermore, results from Edwards and Harrison (1993) required modifying or abandoning many of the substantive conclusions drawn by French et al. (1982), thereby altering the theoretical implications of the study.

## Is Anything Lost by Abandoning Difference Scores?

Despite the apparent advantages of the regression procedure, Tisak and Smith maintain that it is premature to abandon difference scores, arguing that "before we discard this (potentially) theoretically rich concept, more complex difference score functions should be investigated." This apparently reflects the

assumption that, by abandoning difference scores, we are unable to examine theoretical questions of congruence. This assumption is mistaken. As the preceding discussion has shown, the constrained regression equations represented by difference scores are special cases of the unconstrained equations described by Edwards (1994), and any theoretically meaningful functional form depicted by the former can be fully represented by the latter. Furthermore, the unconstrained equations can depict an extensive variety of theoretically meaningful functional forms that difference scores simply cannot represent. Thus, rather than discarding the concept of congruence, the regression procedure permits more rigorous and comprehensive tests of congruence hypotheses while avoiding various problems with difference scores that have plagued this area of investigation for decades.