

RESEARCH NOTES AND COMMENTARIES

MEASURING AND TESTING CHANGE IN STRATEGIC MANAGEMENT RESEARCH

DONALD D. BERGH^{1*} and JAMES F. FAIRBANK²

¹ Smeal College of Business Administration, The Pennsylvania State University, University Park, Pennsylvania, U.S.A.

² College of Business and Economics, West Virginia University, Morgantown, West Virginia, U.S.A.

This paper reviews some critical issues associated with measuring and testing change and then reports on how strategy researchers have addressed those matters. We first discuss three key methodological requirements: reliability assumptions of change variables, correlations between the change variable and its initial measure, and selection of unbiased measurement alternatives. Next, we present data from a content analysis of 126 change studies which suggest that strategy researchers tend not to recognize those requirements. Indeed, the typical approach used to measure and test change (as a simple difference between two measures of the same variable) is usually inappropriate and could lead to inaccurate findings and flawed conclusions. We conclude by offering suggestions for how change can be studied more rigorously. Copyright © 2002 John Wiley & Sons, Ltd.

Change is a complex subject. On one hand, change is intuitive and easy to grasp. Strategists might conceive change as how strategies become different over time, and what effects such differences might have for performance. On the other hand, measuring and testing change involves issues that are not at all obvious. Indeed, researchers need to recognize some very strict and rather obscure methodological requirements that if left ignored can lead to statistical errors and result in incorrect conclusions (Cronbach and Furby, 1970; Linn and Slinde, 1977; Lord, 1963).

Our paper reviews some of the more predominant methodological premises of change and then reports on how strategic management researchers typically apply those requirements. We find that most strategy researchers do not incorporate the critical requirements associated with change. This finding suggests that strategy researchers need to know what the methodological requirements of change are, understand what can go wrong when they study change, and realize what they should do to provide accurate findings and conclusions.

Key words: change scores; repeated measures; methodological critiques

*Correspondence to: Donald D. Bergh, Smeal College of Business Administration, The Pennsylvania State University, 403 Beam Administration Building, University Park, PA 16802, U.S.A.

REQUIREMENTS FOR MEASURING AND TESTING CHANGE

The most basic approach to conceiving and measuring change is as a simple difference between

multiple measures of the same variable. From an operational perspective, change is most directly defined as $C_x = X_1 - X_2$, whereby C_x is called a 'change score' and variable X is measured at time period 1 (X_1) and time period 2 (X_2). Although intuitive, measuring and testing change is not as straightforward as it might seem and requires that researchers attend to two fundamental issues: (1) satisfying reliability assumptions and (2) removing the correlation between the change score and its initial component measure. These issues influence ultimately how to measure and test change more effectively.

Reliability assumptions

Reliability is 'the consistency or stability of a measure or test from one use to the next' (Vogt, 1993: 195). It is a potential problem for measuring and testing change, as the reliability of change scores 'tends to be less than the average reliability of its component parts' (Johns, 1981: 447; Cronbach and Furby, 1970; Lord, 1956, 1963). The basis of this argument is that the change score reflects the combined measurement errors of its component measures, resulting in a comparatively lower reliability value (Bereiter, 1963; Cronbach and Furby, 1970; Lord, 1963).

In addition, the reliability of the change score (C_x) is influenced by the correlation between the component measures (X_1, X_2), which is usually quite high. Even change scores based on highly reliable component variables will have low reliability if those components are correlated highly. For example, when component variables have reliabilities of 0.80, then the change score reliabilities will be 0.60, 0.50, 0.33, and 0.00 when the X_1 and X_2 correlation is 0.5, 0.6, 0.7, and 0.8 respectively (Linn, 1981: 87). These problems have led some to argue that change scores should be abandoned (Cronbach and Furby, 1970; Linn and Slinde, 1977; Edwards, 1994a). Table 1 reports how the reliability of the simple difference change score is influenced by correlations between, and the reliability of, X_1 and X_2 .

However, some suggest that change scores may not be so unreliable after all (Gottman and Rushe, 1993: 970; cf. Allison, 1990). Proponents of this position argue that the primary assumptions underlying the computation of reliability are too restrictive and unrealistic (Rogosa and Willett, 1983;

Table 1. Reliability of simple difference change scores as a product of the reliability of their components and intercorrelation^a

Correlation of X_1 and X_2	Reliability of X_1 and X_2 (assumed to be equal)		
	0.7	0.8	0.9
0.5	0.40	0.60	0.80
0.6	0.25	0.50	0.75
0.7	0.00	0.33	0.67
0.8	—	0.00	0.50
0.9	—	—	0.00

^a Adapted from Linn and Slinde (1977). Formula for computing reliability of simple difference change score is also reported. Requires that component measures have equal reliability, equal variances and a negative correlation between the change score and the initial component measure. See Rogosa (1988) for an alternative viewpoint. Note: A simple difference change score for the variable x is computed as: $C_x = X_1 - X_2$, whereby variable x is measured at time period 1 (X_1) and time period 2 (X_2). Example: The reliability for a simple difference change score is 0.4 when its component measures have reliabilities of 0.7 and a correlation of 0.5.

Zimmerman, 1994). Specifically, traditional approaches for computing change score reliability assume that component measures have equal reliability, equal variances, and a negative correlation between change and initial component measure. However, when equality assumptions and negative correlations are relaxed, reliability of the change score improves markedly (Zimmerman and Williams, 1982). Rogosa shows that when reliabilities of components vary and the correlation between the change score and the initial component measure is set to zero, the change score 'does extremely well' and for a moderate correlation the change score 'is more reliable than the average reliability of the measures. Even for a high correlation, the difference score does rather well compared to the reliability of X ' (Rogosa, 1988: 179).

Correlation with initial measure

Change scores (C_x) are usually correlated with their initial measurement points (X_1) because the change and initial component share variance. This correlation is usually negative because (1) the variance at X_2 is typically less than the variance at X_1 and (2) the correlation between X_1 and X_2 is less than one (Bereiter, 1963). The correlation will be nonnegative when 'the standard deviation of the

postmeasure, X_2 , is larger than that of the premeasure X_1 , and generally substantially so' (Linn and Slinde, 1977: 122; see also this reference for the formula for computing negative correlation).

The correlation between C_x and X_1 is a potential problem for change scores. First, it can lead to regression toward the mean effects (Bohrnstedt, 1969; Cohen and Cohen, 1975). For example, a negative correlation implies that subjects with large positive change scores are more likely to have low scores at X_1 , whereas subjects with high scores at X_1 tend to have low scores at X_2 (Linn and Slinde, 1977). Allison notes that '[b]ecause of the almost universal phenomenon of regression toward the mean from pretest to posttest measurements, $[X_1]$ will usually be negatively correlated with $[X_1 - X_2]$. Thus, individuals with high pretest scores will tend to move down on the posttest, while individuals with low pretest scores will tend to move up' (Allison, 1990: 95). Second, the correlation can lead to interpretation problems (Labouvie, 1982; Nesselroade, Stigler, and Baltes, 1980). Linn and Slinde (1977: 122) noted that larger positive changes 'are more likely to be observed for persons with low X [X_1] scores, whereas persons with high X scores would have large positive [change] only rarely ... if individuals with high [C_x] scores are to be selected, there will be an over-representation of people with low X scores as an artifact due to the negative correlation between [C_x] and X .' These empirical tendencies might lead researchers to make false conclusions about their data; they could easily attribute a finding to some hypothesized effect when it is actually due to regression effects caused by the correlation between C_x and X_1 . Because of these implications, such correlation has been called 'the heart of the problem with simple change scores' (Cohen and Cohen, 1975: 380).

Others, however, have disagreed about the effects of correlations between C_x and X_1 (Humphreys and Drasgow, 1989; Overall, 1989; Overall and Woodward, 1975, 1976). Rogosa, Brandt, and Zimowski (1982) argued that the correlation between the change score and initial measurement arises because of measurement errors and that the correlation does not represent a valid threat to analysis or empirical results. Moreover, Rogosa and Willett observed that 'errors of measurement may produce serious negative bias in the correlation between observed change and observed initial status' (Rogosa and Willett, 1985:

210). Similarly, Rogosa (1988) argued that the correlation between change and initial status depends on the time at which the initial status is measured. Finally, Allison (1990) concluded that regression toward the mean effects depends on context; in particular, such effects do not extend to research designs employing either a nonequivalent control group structure or when two or more stable groups are compared.

Analytical alternatives

The foregoing assessment of reliability and correlations is critical for determining how best to measure and test change. In other words, how change is measured and tested should depend on the reliability and correlational characteristics of the change score components.

Simple difference approach

This alternative for measuring and testing change ($C_x = X_1 - X_2$) can be used only when the component variables have high reliability, low correlation (see Table 1), and when variances are unequal (Rogosa, 1988). When any of these conditions are not met, the simple difference approach is likely to produce biased results and problematic conclusions. Fortunately, alternative analytical techniques have been developed for the settings when the simple difference approach cannot be used (See Table 2).

Residual scores

One way to measure and test change is to compute it as a residual variable. This method may solve the problems caused by the correlation between the change and pretest scores (Linn and Slinde, 1977). In this instance, change is computed as the residual score, R , and regression analysis is used to account for the correlation between C_x and X_1 (Bohrnstedt, 1969; Cronbach and Furby, 1970). The residualized change score ' R is obtained by subtracting the predicted posttest score, Y_2 , from the corresponding observed posttest score, X_2 . The predicted posttest score is obtained from the linear regression of Y [X_2] on the pretest X [X_1]. The zero correlation between X and R follows immediately from the way in which R is derived and is seen as a major advantage over difference scores because residuals do not give an advantage

Table 2. Overview of alternative methods of measuring and testing change

Method	When used	Strengths	Drawbacks
1. Simple difference	High reliability, low correlations	Intuitive, conceptually simple	Vulnerability to errors
2. Residuals	Predicted change is focus rather than actual change	Partial correlation between C and X_1	Not measuring actual change
3. True change	Conceptual change is in simple difference and C and X_1 are correlated highly	Partial correlation, conceptually intuitive	Requires high reliabilities and large n
4. Components	Testing components, not a change score	Overcomes problems of simple difference approach	May not align with some conceptual logics
5. Growth curves	Mapping change	Form and type of change	Not applicable to two-wave data

to persons with certain values of the pretest scores whereas difference scores do' (Linn and Slinde, 1977: 125).

Although the residual change score may reduce the effects of correlations between the component variables (removing the correlation between C_x and X_1), this procedure has some limitations, namely that 'residualized change scores are not in fact measures of change ... [r]ather they are simply measures of whether a person's posttest score is larger or smaller than the value predicted ... [t]o call them growth measures or change scores only tends to confuse the issue' (Linn, 1981: 88; cf. Rogosa, 1988). In addition, the change score may be unreliable (O'Conner, 1972; see Linn and Slinde, 1977, for the formula for computing the reliability of residual change score).

Component scores

Another approach posits that the component measures (i.e., X_1 , X_2), rather than the change score C_x should be the primary focus of the analysis (Linn and Slinde, 1977). Advocates argue that change scores are generally less reliable and less valid than component measures and that change is not a distinct concept by itself (Edwards, 1994a, 1994b; Johns, 1981). These procedures generally depict the component measures as independent variables and outcome (Z) measures as dependent variables. Alternatively, the component measures can be depicted as dependent variables, each of which are regressed separately onto the independent variables, with comparisons then made

between the regression models (recommended statistical tests for these comparisons are discussed below). This alternative has been shown to avoid the reliability concerns associated with algebraic or simple change scores, minimize the potential for confounds of the effects of the component measures, and also provide direct tests of the components relative to the outcome variable. Despite these advantages, some have argued that it is not a direct test of change (Tisak and Smith, 1994).

Growth curves

In circumstances where researchers have access to three or more 'waves' of data, instead of two, growth curves provide a viable analytical alternative. In general, a growth curve is a 'model that represents the level of a variable as a function of time' (Linn, 1981: 89). The slope of the growth curve is derived from the rate of change, and the simplest type of growth model assumes that the level of the variable of interest is a linear function of time (Rogosa *et al.*, 1982). This slope can be estimated by the difference between the posttest and the pretest divided by the difference in time points (Linn, 1981). In addition, growth curves provide the potential for depicting the form of change (Rogosa, 1988; Rogosa and Willett, 1985). Rogosa notes that '[r]esearch questions about growth, development, learning, and the like center on systematic change in an attribute over time, and thus the individual growth curves are the natural foundation for modeling the longitudinal data' (Rogosa, 1988: 172).

In sum, prior research suggests that the adoption of the simple difference method of measuring and testing change should be used only when reliability of the change score is high and no correlation exists between the change score and its initial measure. However, such conditions may not exist in strategic management research and researchers might need to adopt other methods. The potential for problems caused by using inappropriate methods of measuring and testing change raises questions for how strategy researchers have assessed change. Have strategy researchers examined the reliabilities of and correlations among their component variables? How have they measured and tested change?

CHANGE SCORE ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH

To assess whether strategy researchers have recognized the special assumptions underlying change, we conducted a content analysis of change score studies appearing in the *Strategic Management Journal*. Every article appearing from 1985 through 1999 was reviewed to identify studies that used change scores. To be classified as a change score study, a study had to meet two criteria: (1) data were collected at two or more points in time and (2) the analysis involved a comparison of the data between or among the periods represented (studies which pooled data over time were excluded; see Bergh and Holbein, 1997, for a consideration of those studies). An exhaustive review revealed 126 studies that met those criteria.

Each of these studies was analyzed relative to the requirements discussed above, namely: did the authors report change score reliability, correlations between change and initial status, and how was change measured and tested? This evaluation provided several insights into how strategy researchers study change. First, most studies measured and tested change using two waves of data. Change was most frequently computed as simple differences; trend analysis was used the next most frequently, and regression-based techniques were used in less than 1 percent of the total number of articles examined (growth curves did not appear because the studies reported two waves of data only; growth curves can be used when three or more waves of data are examined). Second, very

few of the studies recognized reliability (6 of 126) assumptions and even fewer (4 of 126) controlled for the effects of violating those assumptions. Third, only one study recognized the correlations between the change score components, the correlation of the change score with initial measures, or the implications of selecting any particular analytical alternatives. Two studies acknowledged—but did not test for—the possibility that regression toward the mean effects might threaten the validity of their analytical results.

DISCUSSION

Change is more complicated than might be expected. From a conceptual perspective, change seems quite simple and straightforward, and would seem to reflect a difference between measures or observations. However, from a methodological context, change is fraught with complexities that are far from simple or intuitive, and if ignored can significantly harm empirical results. These tensions suggest that it is important to consider how strategy researchers have studied change. While no one study can address every nuance and subtlety of change, we believe that addressing some of the more substantive issues is a good place to start. Such a review can provide an initial assessment of where the field is currently at with respect to measuring and testing change and thereby provide some signals as to how future research might need to proceed.

The results of our review reveal that the methodological requirements associated with measuring and testing change have, by and large, not been adopted by strategy researchers. This result is easy to understand; researchers would naturally measure and test change in the same way they conceptualize it. However, they may not realize that the most popular approach for computing change, the simple difference method, will produce flawed findings when the component measures are correlated and have equal variances—conditions that are common in strategic management research (Bergh, 1995). Researchers need to recognize that change has a hidden danger. Even in settings whereby component measures may have impressive reliabilities, few variables are perfect and, as a result, the reliabilities of the change scores can easily become zero (Table 1). In addition, regression-toward-the-mean effects can result in plausible alternative

explanations of empirical findings. Strategy researchers are urged to proceed cautiously when measuring and testing change.

When measuring and testing change, we advise researchers to screen their data in terms of reliability of component measures, correlation between the component variables, equality of component variable variances, and correlation with the initial component variable. The results of those tests influence how researchers may best measure and test change (see Table 2). We also suggest that researchers report these statistics, as well as the reliability of their change scores if they use simple difference change scores. Such reporting will enable others to understand the rationale for adopting a particular approach while simultaneously signaling the need to assess data similarly when evaluating change. Moreover, failure to provide such disclosures could cast doubt on studies which measure and test change using the simple difference score.

Our data suggest that most strategy researchers will need to adopt the component approach when measuring and testing change. In doing so, researchers need to recognize how this alternative differs from the simple difference method in both conceptual and empirical ways. We highlight these distinctions with an example.

Example

Consider that a recent line of research has examined whether one reason for the corporate restructurings in the 1980s and 1990s was the increasing concentration of ownership in the hands of 'block holders,' or owners that controlled large blocks of a company's outstanding common stockholdings. The basic argument is that owners that controlled large blocks of stock could influence managers to structure diversification strategies that more closely met their goals of profit maximization. It is believed that these goals and objectives might be attained by refocusing the companies from high diversification to more focused diversification. These restructurings normally occurred by shedding unrelated and noncore business lines, leaving a more narrowly focused firm afterwards. In contrast, when stockholdings are not concentrated, the influence of the large blockholders is reduced. Under these conditions, managers are more capable of pursuing a diversified strategy, as these strategies more clearly meet their goals

and objectives of balancing portfolio and employment risk.

This logic provides a basis for comparing the simple difference and component measure approaches to conceptualizing, measuring, and testing change. The general proposition is that stockholdings held by blockholders are related positively to focused diversification strategies. The simple difference approach to conceptualizing change would produce this hypothesis: Stockholdings held by blockholders are related positively to reductions in diversification. From the component measures approach, the hypothesis is: Stockholdings held by blockholders will be related more negatively to diversification in a subsequent year than in an earlier year.

Explication

Assume that data on blockholdings are collected at time period 1 and the data for diversification are collected for time periods 1 and 2. For the simple difference approach, the hypothesis is supported when blockholdings are associated with positive differences in diversification (diversification in time period 1 is greater than diversification in time period 2). Operationally, the researcher would regress the difference in diversification ($X_1 - X_2$) onto blockholdings in time period 1.

For the component measures approach, the hypothesis reflects the relationship between blockholder stockholdings at time 1 and then at time 2. Blockholder stockholdings would be related negatively to diversification at time period 1 and more negatively at time period 2, indicating that the diversification is lower over the time period of the study, and that both diversification measures are associated with blockholdings in a manner consistent with theoretical prediction. Operationally, the researcher would separately regress diversification at time period 1 and time period 2 onto blockholdings at time period 1. The regression coefficients of these two models would be compared to see if they reflected the hypothesized directions and whether their difference is zero. To support the hypothesis, the coefficient for the blockholding variable in time period 2 would be more negative than the counterpart variable in time period 1, and the difference between the two would be less than zero. The differences in coefficients could be tested using *z*-tests (regression coefficients can be expressed in *z*-scores) or Hotelling's *t*-statistic. We prefer

Hotelling's t , as it tests for differences when the samples are nonindependent, which is clearly the case when change is concerned.

This example indicates that measuring and testing change is not a 'one size fits all' proposition and that sometimes the best approach may be to focus on the component measures rather than the difference between the measures. This component approach is not initially intuitive, but once considered carefully it still captures change in a direct manner while avoiding the problems of the simple difference approach.

In conclusion, we hope that our paper encourages strategy researchers to recognize the complexities of measuring and testing change. By knowing what the requirements are, by understanding what can go wrong, and by knowing how to side-step those problems, strategy researchers will improve research rigor and lead ultimately to stronger empirical foundations for theory development.

ACKNOWLEDGEMENTS

The authors thank Jeffrey Edwards, Larry Williams, David Day, Robert Linn, and John Mathieu for their helpful comments on earlier drafts of this manuscript.

REFERENCES

- Allison PD. 1990. Change scores as dependent variables in regression analysis. In *Sociological Methodology*, Clogg CC (ed.). American Sociological Association: Washington, DC; 93–114.
- Bereiter C. 1963. Some persisting dilemmas in the measurement of change. In *Problems in Measuring Change*, Harris CW (ed.). University of Wisconsin Press: Madison, WI; 3–20.
- Bergh DD. 1995. Problems with repeated measures analysis: demonstration with a study of the diversification and performance relationship. *Academy of Management Journal* **38**: 1692–1708.
- Bergh DD, Holbein G. 1997. Assessing and redirecting longitudinal analysis: demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal* **18**(7): 557–571.
- Bohrnstedt GW. 1969. Observations on the measurement of change. In *Sociological Methodology*, Borgatta EF (ed.). Jossey-Bass: San Francisco, CA; 113–133.
- Cohen J, Cohen P. 1975. *Applied Multivariate Regression/Correlation Analysis for the Behavioral Sciences*. Erlbaum: Hillsdale, NJ.
- Cronbach LJ, Furby L. 1970. How should we measure 'change'—or should we? *Psychological Bulletin* **74**: 68–80.
- Edwards JR. 1994a. Regression analysis as an alternative to difference scores. *Journal of Management* **20**: 683–690.
- Edwards JR. 1994b. The study of congruence in organizational behavior research: critique and a proposed alternative. *Organizational Behavior and Human Decision Processes* **58**: 51–100.
- Gottman JM, Rushe RH. 1993. The analysis of change: issues, fallacies, and new ideas. *Journal of Consulting and Clinical Psychology* **61**: 907–910.
- Humphreys LG, Drasgow F. 1989. Paradoxes, contradictions, and illusions. *Applied Psychological Measurement* **13**: 429–431.
- Johns G. 1981. Difference score measures of organizational behavior variables: a critique. *Organizational Behavior and Human Performance* **27**: 443–463.
- Labouvie EW. 1982. The concept of change and regression toward the mean. *Psychological Bulletin* **92**: 251–257.
- Linn RL. 1981. Measuring pretest–posttest performance changes. In *Educational Evaluation Methodology: The State of the Art*, Berk R (ed.). Johns Hopkins University Press: Baltimore, MD; 84–109.
- Linn RL, Slinde JA. 1977. The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research* **47**: 121–150.
- Lord FM. 1956. The measurement of growth. *Educational and Psychological Measurement* **16**: 421–437.
- Lord FM. 1963. Elementary models for measuring change. In *Problems in Measuring Change*, Harris CW (ed.). University of Wisconsin Press: Madison, WI; 21–38.
- Nesselroade JR, Stigler SM, Baltes PB. 1980. Regression toward the mean and the study of change. *Psychological Bulletin* **88**: 622–637.
- O'Conner EF Jr. 1972. Extending classical test theory to the measurement of change. *Review of Educational Research* **42**: 73–97.
- Overall JE. 1989. Distinguishing between measurements and dependent variables. *Applied Psychological Measurement* **13**: 423–433.
- Overall JE, Woodward JA. 1975. Unreliability of difference scores: a paradox for the measurement of change. *Psychological Bulletin* **82**: 85–86.
- Overall JE, Woodward JA. 1976. Reassertion of the paradoxical power of tests of significance based on unreliable difference scores. *Psychological Bulletin* **83**: 776–777.
- Rogosa D. 1988. Myths about longitudinal research. In *Methodological Issues in Aging Research*, Schaie KW, Campbell RT, Meredith W, Rawlings SC (eds). Springer: New York; 171–209.
- Rogosa D, Brandt D, Zimowski M. 1982. A growth curve approach to the measurement of change. *Psychological Bulletin* **92**: 726–748.
- Rogosa D, Willett JB. 1983. Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement* **20**: 335–343.
- Rogosa D, Willett JB. 1985. Understanding correlates of change by modeling individual differences in growth. *Psychometrika* **50**: 203–228.

- Tisak J, Smith CS. 1994. Defending and extending difference score methods. *Journal of Management* **20**: 675–682.
- Vogt WP. 1993. *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences*. Sage: Newbury Park, CA.
- Zimmerman DW. 1994. A note on interpretation of formulas for the reliability of differences. *Journal of Educational Measurement* **31**: 143–147.
- Zimmerman DW, Williams RH. 1982. The relative error magnitude in three measures of change. *Psychometrika* **47**: 141–147.