LONGITUDINAL STUDIES OF EXPECTANCY THEORY:

SOME STATISTICAL-ANALYTICAL CONCERNS

Working Paper No. 248

L. Delf Dodge

V. Jean Ramsey

L. Delf Dodge
Assistant Professor of Organizational Behavior
Graduate School of Business Administration
The University of Michigan
Ann Arbor, Michigan 48109
(313) 764-1376

V. Jean Ramsey
Assistant Professor of Management
Department of Management
Western Michigan University
Kalamazoo, Michigan 49008
(616) 383-6087

Biographical Notes

L. Delf Dodge is an assistant professor of organizational behavior at the Graduate School of Business Administration, The University of Michigan. She received her M.B.A. (1975) and Ph.D. (1978) in organizational behavior and design from Michigan State University. Her research to date has focused on methodological issues in expectancy models of work effort, developing models of organization design, and theoretical and empirical problems in the definition and measurement variables relevant to organization design.

V. Jean Ramsey is an assistant professor in the Department of Management at Western Michigan University. Her research interests include methodological and measurement problems in organizational research, as well as longitudinal studies of expectancy theory. She received her bachelor's degree from New Mexico State University and an M.B.A. and Ph.D. from The University of Michigan.

Longitudinal Studies of Expectancy Theory:
Some Statistical-Analytical Concerns

ABSTRACT

Difficulties encountered in attempts to measure temporal change in in-
strumentalities and valences in expectancy models of work effort are examined.
Traditionally, correlation analysis has been used. It is argued that the
integrity of individual-level data is best preserved through the use of either
Wilcoxon or sign tests, corrected for random error.

Since their beginnings in the work of Tolman (1932) and Lewis (1935), expectancy theories of motivation have continued to be widely discussed and researched. Mitchell (1974) suggested that the research has been of two general types: studies which attempt to predict levels of work effort and studies which attempt to predict satisfaction with a given choice.

A commonly used version of the effort model (Mitchell, 1974) is represented by the formula

$$W = E \left(\sum_{j=1}^{n} I_{ij} V_j \right),$$

where   W = motivational force;

   E = expectancy that effort leads to performance;

   I_{ij} = instrumentality of performance (first-level outcome (i)

       for the attainment of second-level outcomes (j);

   V_j = valence of second-level outcomes; and

   n = number of outcomes.

Motivational force is a monotonically increasing function of individual valences and instrumentalities associated with anticipated outcomes and expectancies of performance at given levels (Vroom, 1964).

Expectancy theory is recognized as a temporal model. It takes time for the motivational force to develop: a change in the independent variables may not lead to instantaneous shifts in the dependent variables (Kopelman and Thompson, 1976). And yet, the bulk of the research on expectancy models of motivation has used cross-sectional data, which prohibits the examination of changes in individual motivation over time.

By 1974, only three studies had tested models using longitudinal data (Lawler, 1968; Lawler and Suttle, 1973; Sheridan, Slocum, and Richards, 1974). In the next three years, the results of three other longitudinal studies were reported (Kopelman and Thompson, 1976; Sobel and McGuire, 1977; Kopelman, 1977). These studies examined the predictive power of various valence, expectancy, and instrumentality combinations. Few investigated the individual components of the model in depth, even though it has been suggested that mixed empirical support may result from applying the full model in a work setting without adequately investigating the nature of the individual components of the model. Campbell and Pritchard (1976) suggested that instead of conducting large-scale studies with superficial measures of poorly understood variables, researchers should first devote time to a careful analysis of each of the models' several individual components.

Studies which examined individual components of the model over time did so in the context of the stability of responses, using test-retest reliabilities (e.g., Schwab and Dyer, 1973). Connolly (1976) proposed that attempts to predict work effort from measurements of expectancies and valences require strong assumptions concerning the time stability of these characteristics. This suggests that the predictive power of expectancy models may suffer because of perceptual instability in valences, expectancies, and instrumentalities (Sheridan, Slocum, and Richards, 1974). Yet, the dynamic nature of the model implies that changes in valences, expectancies, and instrumentalities accompany changing levels of work effort.

Schwab and Dyer (1973) recognize that low test-retest reliabilities could indicate changes in either motivational levels of the respondents or instrument unreliability. This paradox was noted much earlier.

> When scores on a test are observed to change, how can one tell whether it is the persons who have changed or the tests? If the correlation between pretest and post-test is reasonably high, we are inclined to ascribe change scores to changes in individuals. But if the correlation is low, we may suspect that the test does not measure the same thing on the two occasions (Bereiter, 1963, p. 11).

We are, as yet, incapable of distinguishing between the two.

Studies are needed to investigate each of the major components (expectancies, instrumentalities, and valences), as well as those variables thought to moderate the relationships among the major components (e.g., equity, role perceptions), and the changes which may occur in them over time. The results of one such study, investigating two of the several components, are reported here. Specifically, this research examines whether instrumentalities and valences change, and, if so, whether such changes appear simultaneously or prior to changes in effort levels.

THE STUDY

The study was conducted in five sections of an introductory organizational behavior course at a large midwestern university and included 191 students in the third year of their undergraduate business curriculum. To determine relevant outcomes, the first questionnaire, administered during the first week of the semester, included a list of possible outcomes generated by the researchers which students were asked to rank order. Those outcomes ranked highest were a feeling of having learned something, a feeling of achievement, a high grade, a sense of personal growth, acquisition of new knowledge and skills, the grade they expected, a sense of having met a challenge, and earning credit toward a degree. These items were included in all subsequent questionnaires.

In the class session before each examination, students were asked to indicate, on a scale of one (extremely likely) to seven (extremely unlikely), the probability that each outcome would follow from their performance

(instrumentality). Similarly, subjects were asked to indicate, on a scale of one (extremely important) to seven (extremely unimportant), their perception of the degree of valence of each of the outcomes at that time.

Three examinations were administered over ·a period of four months, at approximately equal intervals of five weeks. As a result, three separate assessments were obtained of valences and instrumentalities associated with the eight outcomes identified by subjects at the beginning of the course. Average response rate for the three administrations was approximately 78 percent.

Standard routines in MIDAS (Michigan Interactive Data Analysis Systems, 1974) and SPSS (Statistical Package for the Social Sciences, 1975) were used to analyze the data. Following the empirical precedent set by earlier studies of expectancy models of work motivation, correlational analyses were used to examine the stability of instrumentalities and valences.

CORRELATIONAL RESULTS

Across-subject correlational results appear in Table 1.

Insert Table 1 about here

In all instances, correlations between values of variables across time periods were significant. One might conclude from these data that the valences and instrumentalities of respondents did not change over the course of the study. The importance an individual assigns to outcomes and the likelihood of outcome occurrence at a given level of performance might be reported as stable. From this, it could be inferred that once employers had determined how important the organization's rewards were to an individual, they could expect those rewards to remain a source of motivation for that employee for some time.

Table 1

Correlations of Valences/Instrumentalities
across Subjects over Time

	TIME 1 - TIME 2	TIME 2 - TIME 3
Valence 1	.26****	.41****
Valence 2	.31****	.41****
Valence 3	.40****	.52****
Valence 4	.16*	.20**
Valence 5	.27****	.42****
Valence 6	.55****	.49****
Valence 7	.56****	.41****
Valence 8	.38****	.52****
MEAN of Valences	.36****	.42****
Instrumentality 1	.08	.40****
Instrumentality 2	.24****	.30****
Instrumentality 3	.34****	.40****
Instrumentality 4	.22****	.29****
Instrumentality 5	.32****	.48****
Instrumentality 6	.43****	.44****
Instrumentality 7	.48****	.35****
Instrumentality 8	.34****	.45****
MEAN of Instrumentalities	.36****	.42****

$* \quad p \leq .10$
$** \quad p \leq .05$
$*** \quad p \leq .01$
$**** \quad p \leq .001$

Since expectancy models were designed to predict individual, not aggregate, behavior, we question whether such conclusions are warranted.

Across-subject correlational analyses such as these pool the data of multiple subjects, calculate the variances of the variables to be correlated, and then check to see if the level of association between the variables is greater than or less than that which we could expect to occur by chance. Under some circumstances, this computational sequence is appropriate, and does not damage the essence of the data upon which the test is used. However, in the study of expectancy theory, the salient unit of analysis is the individual and his or her perceptions of the alternative behaviors and outcomes available at a given point in time. Patterns of valence and instrumentality reports of single individuals are important; however, when data are combined across groups of individuals, results become uninterpretable. We can no longer tell if a person's score is fluctuating or remaining stable, because the correlational procedure has pooled the responses of several persons into a single measure.

It is possible for change to be occurring within samples which show positive, negative, or zero correlations. However, if response means are reported along with the correlations, it becomes possible to detect change. But here again, individual level changes are masked when the responses of several people are pooled in order to calculate the mean. Changes occurring within individual subjects may balance one another, creating the illustion of stability; that is, while one person's scores are changing in an upward direction, another's scores may move down, and still another's may not change at all. Means will report, however, that the measures taken in the first time period are not significantly different from those in the second time period.

Also, correlational procedures assume that the samples involved are independent. When an individual's change over time is examined in the context of expectancy theory, the samples (time periods) involved are not independent.

Finally, correlations tell us about the amount of variation occurring in the samples over several measures, but nothing about specific sources of variance or the type of adjustments being made.

In an attempt to retain the value of individual level data, correlations were run within subjects, then, using r to z transformations, averaged across the subject pool (Snedecor, 1946). Results of these calculations appear in Table 2.

Insert Table 2 about here

The correlations within subjects (though not statistically significant because of the size of n) are substantially greater than those reported across subjects. This corroborates the earlier results which indicated little change in instrumentalities and valences over time.

Using within-subject correlations solves some of the problems encountered in across-subject analyses, but creates others. We do gain more useful motivational information about each individual, but we are forced to pool several of that individual's responses. When within-subject analyses are performed, n, for statistical purposes, is effectively reduced to one. To calculate correlations, some variance is needed. Unfortunately, there is no variance in a sample consisting of a single observation. To introduce variance, multiple measures of a single variable or clusters of similar variables may be entered into the equation, but then we no longer know what is happening within the variable clusters for that individual. Interpretive ability is thus diminished.

Table 2

Average Correlations* of Valences/Instrumentalities
within Subjects over Time

VALENCES

TIME 1 - TIME 2   TIME 2 - TIME 3

.60   .63

INSTRUMENTALITIES

TIME 1 - TIME 2   TIME 2 - TIME 3

.68   .73

*r to z transformed

Additionally, the n for the analysis will likely remain so low (eight, in this study) that a correlation of approximately .71 is necessary to achieve even marginal levels of significance (p = .05, n = 8). As n decreases, the problem intensifies (for p = .05, n = 5, r = .87; for n = 3, at p = .05, r = .99). Differences in sets of scores for an individual respondent cannot be detected by a correlation, because n is too low for the test to be sensitive to change.

There is also some question about what a correlation can tell us. Even if a correlation is high between time periods, we cannot confidently infer that there has been no change; we can only conclude that the variables measured at time one and at time two are varying together. For example, if the variables are increasing or decreasing at similar rates, they are both changing, but would be highly correlated.

Average correlations of within-subject responses, then, do not appear to offer a viable solution to measuring change within an expectancy model. Other avenues need to be explored.

RAW SIGN TEST RESULTS

From the same data, sign tests give us information similar to correlations, but yield more specific information about changes in a single subject's score. Sign tests are applicable to cases in which a researcher wishes to establish levels of similarity between two conditions (Siegel, 1956). The test requires that the underlying distribution of the variables involved be continuous, and that subjects for comparison be matched with respect to relevant extraneous variables. In within-subject analyses, subjects function as their own control, i.e., comparisons over time are of a single individual's responses to a variable at several points in time.

Contrary to the correlational results reported in Tables 1 and 2, examination of Table 3 (first 2 columns of each time comparison) reveals that there was, indeed, activity between time periods.

Insert Table 3 about here

Subjects' responses changed in both directions. As many as 62/118 (52 percent) of the responding sample reported different valences or instrumentalities between two points of measurement (the instrumentality of outcome 3 between times 2 and 3). It is interesting to note that the correlation for this same variable, over the same time period, was .40 ($p < .001$). At a significant level of correlation, it might seem surprising to find 52 percent of the sample's respondents changing their scores between measures.

The raw sign test tells how many subjects have changed their rating, and whether they made their adjustments in an upward or downward direction. Raw sign tests do not, however, take the magnitude of the changes into account, nor do they allow for random error which may have occurred.

The more powerful Wilcoxon matched-pairs signed-ranks tests were used on the study data to determine whether the magnitude of change in subjects' reported valences and instrumentalities affected results obtained through the sign tests (Table 4).

Insert Table 4 about here

The two sets of tests produced comparable results. Variables which were shown to be significantly different between time periods using the sign test were also shown to be significantly different using the Wilcoxon test. There was essentially no difference between the two sets of results.

Table 3

Sign Tests of Valences/Instrumentalities over Time

	TIME 1 - TIME 2				TIME 2 - TIME 3			
	-DIFF	+DIFF	Z	P(Z)*	-DIFF	+DIFF	Z	P(Z)*
Valence 1	15	37	2.9	.003	13	29	2.31	.020
Valence 2	18	40	2.76	.005	27	21	0.72	.471
Valence 3	24	35	1.32	.186	25	25	0.14	.888
Valence 4	11	38	3.71	.000	24	19	0.61	.541
Valence 5	32	30	0.12	.894	25	32	0.79	.429
Valence 6	18	15	0.35	.718	19	19	0.16	.872
Valence 7	21	25	0.44	.660	24	21	0.29	.771
Valence 8	21	25	0.44	.660	36	23	1.56	.118
Instrumentality 1	29	32	0.25	.802	20	28	1.01	.312
Instrumentality 2	21	36	1.85	.064	29	23	0.69	.490
Instrumentality 3	30	24	0.68	.496	31	31	0.12	.303
Instrumentality 4	18	33	1.96	.050	30	22	0.97	.332
Instrumentality 5	29	28	1.22	.222	26	29	0.27	.787
Instrumentality 6	32	23	1.08	.2802	31	28	0.26	.794
Instrumentality 7	21	38	2.08	.037	39	20	2.34	.019
Instrumentality 8	24	20	0.45	.752	36	17	2.47	.013

*2-tailed

Table 4

Matched-Pairs Signed-Ranks (Wilcoxon) Tests
of Valences/Instrumentalities over Time

	TIME 1 - TIME 2			TIME 2 - TIME 3		
	-DIFF	+DIFF	P	-DIFF	+DIFF	P
Valence 1	27.23	26.20	.01	17.00	23.52	.004
Valence 2	25.00	31.52	.002	26.83	21.50	
Valence 3	28.46	31.06		24.50	26.50	
Valence 4	26.73	24.50	.002	21.00	23.26	
Valence 5	32.91	30.00		25.94	31.39	
Valence 6	16.92	17.10		18.00	21.00	
Valence 7	24.79	22.42		24.25	21.57	
Valence 8	21.86	24.88		25.00	21.17	.08
Instrumentality 1	30.17	31.75		24.60	24.43	
Instrumentality 2	27.67	29.78	.05	27.55	25.17	
Instrumentality 3	25.77	29.67		32.44	30.56	
Instrumentality 4	24.00	27.09	.03	27.47	25.18	
Instrumentality 5	29.45	28.54		28.62	27.45	
Instrumentality 6	26.44	30.17		29.24	30.84	
Instrumentality 7	29.81	30.11	.05	29.27	31.42	.05
Instrumentality 8	22.44	22.57		25.74	29.68	.06

CORRECTED SIGN TEST RESULTS

The raw change or raw gain scores, upon which sign tests are based, have been criticized because they are formed by a simple subtraction of scores. This can lead to fallacious conclusions, primarily because such scores may be systematically related to any random error of measurement (Cronbach and Furby, 1970).

Approximation to a normal curve, the underlying condition assumed by a sign test to exist, is greatly enhanced when a correction for continuity[1] is employed (Siegel, 1956). The correction is effected by reducing the difference between the observed gains (or losses) and the expected number of gains (or losses) for a given sample size. The correction for continuity produces a z-score from which we can determine the probabilities that the change is attributable to true sample change, or to error variance. Thus, individual level data are essentially preserved, and random variation in reported scores is accounted for. Results of sign tests, corrected for continuity, appear in Table 3.

It becomes apparent that the conclusions we might draw from these data are considerably different from those indicated by the correlational results. Between times one and two, six variables changed significantly (outcome valences 1, 2, 4 and outcome instrumentalities 2, 4, 7), while ten did not. Between times two and three, three variables (outcome valence 1 and outcome instrumentalities 7 and 8) changed significantly. In total, nine variables shifted significantly between time periods. Correlations reflected none of these changes.

Sign tests, corrected for continuity, appear to give researchers useful information about changes in individual subjects' responses. With other data sets where magnitude of differences is a factor, the Wilcoxon test may be more

appropriate. Only if there is variation in magnitude of change within the sample would the two tests have different results.

CONCLUSION

Expectancy models of motivation attempt to describe the general process of evaluation used by individuals in choosing a level of work effort to expend on a given task. An individual's history, experience with similar situations, or expectations about a new choice situation will affect his or her behavior in the current situation. When the current situation becomes history, it too will affect future choice. In order to test expectancy models, then, it is essential that the time dimension be taken into account.

The traditional method of testing temporal effects expectancy models of work motivation has been correlation. Inferences about the stability of subject perceptions are made on the basis of correlational results. There is some question, however, whether such procedures are really appropriate.

Across-subject correlational analyses, which combine data across sample subjects, are viewed as inappropriate for most applications. Even in within-subject studies, there are problems in using correlational analyses in tests of expectancy models. The source of the difficulties is in the composition of n, since n becomes the number of measures instead of the number of subjects, effectively reducing the sample size. Rather than pooling individuals as a source of variance in across-subjects analysis, we are forced to cluster variables to achieve variance. As a result, we are forced to cluster variables to achieve variance. As a result, the integrity of specific variables is compromised. In addition, the magnitude of the correlation necessary to achieve significance under such conditions is almost unattainably large.

Sign tests can detect change more effectively, yield more specific information about changes in subjects' scores, and preserve the essence of individual level data. After correcting sign tests for random change, it became evident that variables in this study which correlational results might have reported as "stable" indeed were undergoing considerable change. It would appear advisable to exercise some caution in the choice of methods for testing change in expectancy models, since this choice seems to have a substantial effect on the results obtained.

Although considerably less complex than tests of correlation, sign tests (or Wilcoxon matched-pairs signed-ranks tests) may well be more appropriate for use in testing expectancy models of work motivation. The ability to leave individual-level data relatively intact, and to record the amount, direction, and significance of change, make such tests attractive for this type of application.

There are some problems, however, which have yet to be resolved. One difficulty inherent in testing any variable over time is our inability to completely reject the alternative explanation of change as unreliability in measurement instruments. Such problems in demonstrating change are particularly important in the context of expectancy theory research.

While controversy on this issue will continue, we can progress by carefully examining the assumptions underlying the statistical techniques we use. In addition to exercising care in the choice of analytical modes, we must be cautious in our interpretation of results. Only under such conditions can expectancy models be given a fair test.

FOOTNOTE

1. $Z = \dfrac{(X + .5) - 1/2\ N}{1/2\ N}$

REFERENCES

1. Bereiter, C. "Some Persisting Dilemmas in the Measurement of Change."

    In Chester W. Harris (ed.), Problems in Measuring Change. Madison: The

    University of Wisconsin Press, 1963.

2. Campbell, J. P., and R. D. Pritchard. "Motivation Theory in Industrial

    and Organizational Psychology." In Marvin D. Dunnette (ed.), Handbook of

    Industrial and Organizational Psychology. Chicago: Rand-McNally, 1976.

3. Connolly, T. "Some Conceptual and Methodological Issues in Expectancy

    Models of Work Performance Motivation." Academy of Management Review,

    Vol. 1 (1976): 37-47.

4. Cronbach, L. J., and L Furby. "How Should We Measure 'Change'--Or Should

    We?" Psychological Bulletin, Vol. 74 (1970): 68-80.

5. Fox, D. J., and K. E. Guire. Documentation for MIDAS. Ann Arbor, Mich.,

    Statistical Research Laboratory, The University of Michigan, 1976.

6. Kopelman, R. E. "Across-Individual, Within-Individual and Return on

    Effort Versions of Expectancy Theory." Decision Sciences, Vol. 8 (1977):

    651-662.

7. _____ and P. H. Thompson. "Boundary Conditions for Expectancy Theory

    Predictions of Work Motivation and Job Performance." Academy of Management

    Journal, Vol. 19 (1976): 237-258.

8. Lawler, E. E., III. "A Correlational-Causal Analysis of the Relationship

    between Expectancy Attitudes and Job Performance." Journal of Applied

    Psychology, Vol. 52 (1968): 462-468.

9. _____ and J. L. Suttle. "Expectancy Theory and Job Behavior."

    Organizational Behavior and Human Performance, Vol. 9 (1973): 482-503.

10. Lewin, K. A Dynamic Theory of Personality. New York: McGraw-Hill, 1935.

11. Mitchell, T. R. "Expectancy Models of Job Satisfaction, Occupational Preference and Effort: A Theoretical, Methodological, and Empirical Appraisal." Psychological Bulletin, Vol. 81 (1974): 1053-1077.

12. Nie, N. H.; C. H. Hull; G. Jenkins; K. Steinbrenner; and D. H. Bent. Statistical Package for the Social Sciences, 2nd edition. New York: McGraw-Hill, 1975.

13. Schwab, D. P., and L. D. Dyer. "The Motivational Impact of a Compensation System on Employee Performance." Organizational Behavior and Human Performance, Vol. 9 (1973): 215-225.

14. Sheridan, J. E.; J. W. Slocum, Jr.; and M. D. Richards. "Expectancy Theory as a Lead Indicator of Job Behavior." Decision Sciences, Vol. 5 (1974): 507-523.

15. Siegel, S. Nonparametric Statistis for the Behavioral Sciences. New York: McGraw-Hill, 1956.

16. Snedecor, G. W. Statistical Methods, 4th edition. Ames, Iowa: The Iowa State College Press, 1946.

17. Sobel, R. S., and H. McGuire. "Time Perspective and Longitudinal Tests of a Cognitive Model of Satisfaction." The Journal of General Psychology, Vol. 96 (1977): 177-184.

18. Tolman, E. E. Purposive Behavior in Animals and Men. New York: Century Co., 1932.

19. Vroom, V. H. Work and Motivation. New York: John Wiley & Sons, 1964.