Validity and Reliability of the Confirmation of Expectations Paradigm As a Determinant of Consumer Satisfaction

Ved Prakash, D.B.A.

Florida International University

The purpose of this paper is to describe a measurement problem confronting researchers who would study the confirmation of expectations by means of difference scores. The concept of confirmation of expectations has been used specifically to study consumer satisfaction and, more generally, to understand the bases of buyer behavior (cf. Howard and Sheth 1969; Oliver 1980; Engel and Blackwell 1982). Typically, expectations on a set of dimensions are measured prior to purchase and evaluations are made on the same set of dimensions after the purchase. The difference between these expectations and evaluations represents the construct of confirmation of expectations.

The problem with this approach is not a conceptual one. The notion that, say, satisfaction with purchase derives from confirmed expectations makes good sense on an a priori basis and, in fact, has sound theoretical grounding in Helson's (1959) adaptation level theory and Thibaut and Kelley's (1959) comparison level theory. The problem is one of measurement. To see the problem in a more general perspective, we can turn to the psychometric literature in the field of Psychology. As has been discussed in detail by several authors (e.g., Cronbach and Furby 1970; Lord 1963; Magnusson 1965), anytime the score on one variable is subtracted from the score on another variable to form the score on the variable of interest (the difference score variable), there is the potential threat of low reliability of the differ-

 Journal of the Academy of Marketing Science
Fall, 1984, Vol. 12, No. 4, 063-076
0092-0703/84/1204-0063 $2.00

ence score variable. Consider the following formula given by Lord (1963) for estimating the reliability of a difference score:

$$r_{DD'} = \frac{s_y^2 r_{yy'} - 2r_{xy}s_y s_x + s_x^2 r_{xx'}}{s_y^2 - 2r_{xy}s_y s_x + s_x^2} \qquad (1)$$

Where:

$r_{DD'}$ = reliability of the difference score (y-x)
$r_{xx'}$ = reliability of the first measure (e.g., prepurchase expectations)
$r_{yy'}$ = reliability of the second measure (e.g., postpurchase evaluation)
s_x^2 = variance of the first measure
s_y^2 = variance of the second measure
r_{xy} = correlation between the first measure and the second measure

As can be seen in this formula, the reliability of the difference score decreases as the variance of either measure decreases. The reliability of the difference score measure also decreases as the correlation between the two component measures *increases*. Thus, if one wanted to increase the difference score reliability, one might attempt to *lower* the correlation between, say, a measure of prepurchase expectation and a measure of postpurchase evaluation for the same attribute (or set of attributes). But this would lead to a paradox since, with lower correlation between the two measures one would have less assurance that the same attribute (or set of attributes) was, in fact, being measured. Furthermore, when the component variables are measured at different points in time, difference score unreliability capitalizes on any effects of regression-to the-mean overtime. Such regression effects contribute to a lowering of the difference score reliability. As will be shown in the data presented later in this paper, the empirical estimates of the reliability of difference score measures are at best modest and rather low. The problem that this poses for theory-building and efforts to ascertain the construct validity of confirmation of expectations measures is straightforward. At the very least, low reliability obscures the true validity of the construct; most often it reduces the observed validity of a construct.

EMPIRICAL STUDY

Overview

The data described and discussed below are from a larger study by the author. The product chosen for this study was beer. The sample consisted of 402 students in the College of Business Administration at a large Southeastern university, chosen for their familiarity with the product.

Method

Data were collected in two stages. In the first stage data were collected on three types of expectations: Predictive expectations, i.e. how a brand is *likely* to perform on brand attributes; Normative expectations, i.e. how a brand *should* perform in order for the consumer to be completely satisfied; Comparative expectations, i.e. consumer *expectations from similar other brands*. Data on all these expectations were collected on seven-point bipolar scales on seven attributes of beer such as good taste, pleasant aftertaste, good value for price, not filling, recommendation of friends, and good brand reputation. In the second stage (three weeks after the first) data were collected on postpurchase evaluation on seven point bipolar scales on the same brand attributes. Data were also collected on overall satisfaction with this last purchase experience on a seven point bipolar scale ranging from Extremely Dissatisfied to Extremely Satisfied. At this stage information was also gathered on the intention to repurchase on an eleven-point probability scale ranging from No Chance At All to Absolutely Certain (cf. Juster 1966). Thus, data were collected on four determinants of overall satisfaction. (a) Confirmation of predictive expectations, i.e. the difference between postpurchase evaluation and predictive expectations. (b) Confirmation of Normative Expectations, i.e. the difference between postpurchase evaluation and normative expectations. (c) Confirmation of comparative expectations, i.e. the difference between postpurchase evaluation and comparative expectations. (d) Postpurchase Evaluation by itself can also be considered a determinant of satisfaction (Lingoes and Pfaff 1972); Czepiel and Rosenberg 1976)

Reliability Analysis

The overall confirmation of expectations measures were derived by summing up the differences between postpurchase evaluation and prepurchase expectations on brand attributes. Since multiple difference scores were

combined here in forming the composite confirmation of expectation measures, one simplifying assumption must be made to permit a meaningful application for formula (1): It is assumed here that the sum of individual scores for two variables across a common set of dimensions

$$\left(\text{e.g.} \sum_{i=1}^{n} (A_i - B_i) \right)$$

equals the sum of the first variable across dimensions minus the sum of the second variable across dimensions.

$$\left(\text{e.g.} \left(\sum_{i=1}^{n} A_i \right) - \left(\sum_{i=1}^{n} B_i \right) \right)$$

To use formula (1) for estimating the reliability of the difference scores (the confirmation of expectations measures), information was needed on three types of parameters. First, the reliabilities of prepurchase expectations (and postpurchase evaluations) were estimated by computing Cronbah's Coefficient Alphas (see Magnusson 1965) on the composite scores of attributes ratings. Second, variance estimates were computed on the composite scores across individuals. Finally, the correlation between prepurchase and postpurchase composites was computed for each measure. Then formula (1) was applied.

Table 1 presents the reliability coefficients for the composite measures of the four determinants of overall satisfaction. On the basis of these coefficients the following conclusions can be drawn. First, in general the reliabilities of the confirmation of expectations measures are very low. The confirmation of predictive expectations has a poor reliability of .19. The confirmation of normative expectations and the conformation of comparative expectations also have low reliability coefficients of .48 and .48 respectively. As compared to these low reliabilities, the postpurchase evaluation has a reasonably good reliability coefficient of .71 largely because this measure is not attenuated by the difference scores (the cause of low reliabilities of the confirmation measures). Secondly, as can be seen from Table 2, the low reliabilities of the confirmation measures are also responsible for the low correlations of these measures with overall satisfaction .19 for the confirmation of predictive expectations and .29 for the confirmation of comparative expectations and .41 for the confirmation of normative expectations. In comparison, the postpurchase evaluation has a correlation of .59 with overall satisfaction.

TABLE 1

RELIABILITY COEFFICIENTS FOR VARIOUS DETERMINANTS OF SATISFACTION (n=231)

Determinant Measure	Reliability Coefficient
(a) Confirmation of Predictive Expectations	.19
(b) Confirmation of Normative Expectations	.48
(c) Confirmation of Comparative Expectations	.48
(d) Postpurchase Evaluation	.71

TABLE 2

CORRELATION COEFFICIENTS BETWEEN OVERALL SATISFACTION AND VARIOUS DETERMINANTS OF SATISFACTION (n = 231)

Determinant Measure	Correlation Coefficient
Confirmation of Predictive Expectations	.19*
Confirmation of Normative Expectations	.41*
Confirmation of Comparative Expectations	.29*
Postpurchase Evaluation	.59*

* $P \leq .01$

RESULTS AND DISCUSSION

According to the assumptions of the weak causal order and causal order closure (Kim and Kahout 1975, pp. 383-387), if a sequence of variables is intercorrelated then the correlations between the two preceding variables should be higher than the correlations between the two succeeding variables. In other words, for a sequence of variables to represent consistency, the diagonal correlations should be higher than the correlations in the corresponding rows and columns.

Table 3 shows the intercorrelations between the variables in the satisfaction process with the confirmation of predictive expectations being the focal part of it. The first row shows that the predictive expectations correlate well with postpurchase evaluation with a coefficient of .61 and this coefficient is higher than all other correlations in that row. In the second row, however, the correlation coefficient between the confirmation of predictive expectations and postpurchase evaluation .39 is lower than the correlation coefficient between overall satisfaction and postpurchase evaluation .59; this correlation is also lower than the correlation coefficient between predictive expectations and the confirmation of predictive expectations i.e. -.50. Consequently, this row does not meet the requirements of the causal order consistency. Similar problems exist with regard to the third row and third column, where the diagonal coefficient representing the correlation between the confirmation of predictive expectations and overall satisfaction .19 is lower than the correlation between overall satisfaction and postpurchase evaluation .59. Consequently the requirements of the causal order are not met. It is clear that the confirmation of predictive expectations is the confounding variable in the satisfaction process and if this variable were dropped from the process the diagonal correlations would be consistently higher than the correlations in the corresponding rows and columns.

Table 4 presents similar intercorrelations with regard to the confirmation of normative expectations in the satisfaction process. Here none of the diagonal correlation coefficients meet the requirements of the causal order. In the first row, the correlation between normative expectations and postpurchase evaluation .35 is lower than the correlation between normative expectations and the confirmation of normative expectations i.e. .60. In the second row the correlation between the confirmation of normative expectations and postpurchase evaluation .54 is slightly lower than the correlation between postpurchase evaluation and overall satisfaction .59. This correlation is also lower than the correlation between the confirmation of normative expectations and normative expectations .60. The diagonal

TABLE 3

INTERCORRELATION COEFFICIENTS FOR CONFIRMATION OF PREDICTIVE EXPECTATIONS IN SATISFACTION PROCESS (n = 231)

Item	Post-Purchase Evaluation	Confirmation of Predictive Expectations	Overall Satisfaction	Repurchase
Predictive Expectations	.61*	-.50*	.39*	.13**
Postpurchase Evaluation		.39*	.59*	.30*
Confirmation of Predictive Expectations			.19*	.18*
Overall Satisfaction				.26*

*P ≤ .01
**P ≤ .05

coefficient in the third row also does not meet the requirements of the causal order. The correlation coefficient between the confirmation of normative expectations and overall satisfaction .41 is much lower than the correlation between postpurchase evaluation and overall satisfaction .59. Once again if the confirmation of normative expectations variable were dropped from the satisfaction process the diagonal correlation coefficients would be consistently higher than the correlations in the corresponding rows and columns and the requirements of the causal order would be satisfied.

Table 5 presents the intercorrelations between the variables with the confirmation of comparative expectations in the satisfaction process being the focal part. The analysis shows that the diagonal correlation coefficients are consistently lower than the correlations in the relevant rows and columns. The requirements of the causal order are not met and would be met if the confirmation of comparative expectation variable were dropped from the satisfaction process.

Overall, postpurchase evaluation ratings are providing the best correlations with satisfaction and repurchase. This is largely because of the higher reliability coefficient of postpurchase ratings as compared to the low reliability coefficients of the difference scores from which the three confirmation of expectations measures have been computed. Postpurchase evaluation ratings and overall satisfaction are found to be equally good predictors of the intention to repurchase. Here two issues may be raised regarding the prediction of the intention to repurchase. The first issue may be regarding the fact that the overall satisfaction with the brand was based on the act of purchase of the last brand but the intention to repurchase may have been based upon a series of previous impressions and not just the last purchase. The answer to this objection would be that for *frequently* purchased products such as beer, *expectations* have already been formed because of long consumption experience; consequently, even the experience of overall satisfaction may in reality be based upon a series of previous impressions not just the last purchase. The act of last purchase was only a convenient method of measuring the level of satisfaction with a specific brand. Alternatively, it would have been necessary to undertake an expensive multistage longitudinal study to measure satisfaction and repurchase. The second objection with regard to the prediction of the intention of repurchase may be that once a consumer has selected a brand for purchase then in order to avoid cognitive dissonance he would indicate a high probability of repurchase of the same brand. This undoubtedly is one of the drawbacks of depending upon postpurchase measures to predict repurchase. In the present study, however, such a problem did not exist because the

TABLE 4

INTERCORRELATION COEFFICIENTS FOR CONFIRMATION OF NORMATIVE EXPECTATIONS IN SATISFACTION PROCESS (n = 231)

Item	Post-Purchase Evaluation	Confirmation of Normative Expectations	Overall Satisfaction	Repurchase
Normative Expectations	.35*	-.60*	.11	.04
Postpurchase Evaluation		.54*	.59*	.30*
Confirmation of Normative Expectations			.41*	.22*
Overall Satisfaction				.26*

*P $\leq$ .01

TABLE 5

INTERCORRELATION COEFFICIENTS FOR CONFIRMATION OF COMPARATIVE EXPECTATIONS IN SATISFACTION PROCESS (n = 231)

Item	Post-Purchase Evaluation	Confirmation of Comparative Expectations	Overall Satisfaction	Repurchase
Prepurchase Expectations	.45*	-.53*	.28*	.04
Postpurchase Evaluation		.52*	.59*	.30*
Confirmation of Comparative Expectations			.29*	.25*
Overall Satisfaction				.26*

*P $\leq$ .01

correlation coefficient between post purchase evaluation and the intention to repurchase was relatively low i.e. .30. The intention to repurchase depends upon many factors such as the availability of other brands, price reduction etc. Although consumers are known to be brand loyal with respect to beer yet they do not necessarily stick to one brand all the time. Instead, they may have two or three brands in their evoked set.

CONCLUSIONS

In this paper we have examined some aspects of reliability and validity with regard to the confirmation of expectations paradigm. It has been shown that the confirmation measures based on difference scores have low reliabilities and consequently have low correlations with the criterion measure of overall satisfaction. Also, it has been seen that postpurchase measures (or some direct measures) are good correlates of satisfaction (and repurchase). Consequently, for the marketing manager there may be no need to go through the lengthy procedure of deriving the confirmation of expectations measures.

The major advantage of the confirmation paradigm is for diagnostic applications. If that is the purpose of a research project then it is recommended that some reliable scales be developed to measure the construct of confirmation. Researchers who choose to employ the difference score method should expect to encounter low reliabilities of these measures. Low reliabilities of difference score measures such as .48 and .19 should not be considered acceptable. Low reliability reduces the power of the statistical tests, lowers estimates of effect sizes, and clouds our understanding of the construct validity of theoretically relevant variables.

One suggestion toward the development of more reliable scales is the use of comprehensive scales incorporating the idea of confirmation. For example: "How Well did Brand X Perform on "Attribute A"?

"Much Worse Than Expected"	"Worse Than Expected"	"About the Same as Expected"	"Better than Expected"	"Much Better Than Expected"
1	2	3	4	5

This approach avoids the use of difference scores and has in fact previously been used (Oliver 1980). This approach, however, does suffer from a bias toward the subject in the postpurchase situation rationalizing his or her prepurchase expectations. To help reduce such a bias, we suggest combining this approach with the following procedure: measure and record

prepurchase expectations and then ask the subjects to review these expectations immediately prior to making ratings on the degree to which confirmations have been realized. Such an approach would force the subjects to review their actual prepurchase expectations. This approach could be further improvised by requesting the subjects to justify actual confirmation and disconfirmation of expectations.

Secondly, if at all difference score methods are used, two strategies could be adopted to improve reliability. (1) Increase the reliability of the two component variables by increasing the number of items or by increasing the homogeneity of items (or product attributes) in the component measures. (2) Attempt to increase the number of types of confirmation of expectations measures used to predict satisfaction; even with low reliabilities for each of the individual confirmation measures, given enough of them, prediction can be substantially improved.

Finally, no matter what measurement strategy is employed, researchers in this area should be encouraged to examine empirically the quality of their measures in terms of classical reliability concepts as well as other factors which may impede validity such as range restriction, criterion contamination, and generalizability of measurement procedures across item pools, product types, settings, time intervals, and subject characteristics. A next logical step would be to apply generalizability theory concepts and estimation procedures (cf. Cronbach, Gleser, Nanda, and Rajaratnam 1972) to measures of interest for marketing research. Our hunch is that such endeavors will expose some disappointing data on the generalizability of constructs across measurement facets, but the knowledge gained will hopefully point toward remedial efforts that will ultimately lead to enhanced prediction and more valid models.

REFERENCES

Cronbach, Lee J., and Furby, L. 1970. "How We Should Measure 'Change' or Should We?" *Psychological Bulletin* 74 (July), 68-80.

Cronbach, Lee J., Gleser, Goldine C., Nanda, Harinder and Rajaratnam Negeswari. 1972. *The Dependability of Behavioral Measurements: Theory of Generalization for Scores and Profiles*, New York: John Wiley.

Czepiel, John A., and Rosenberg, Larry, J. 1976. "The Study of Consumer Satisfaction: Addressing the 'So what' Question," in Keith Hunt (ed.) *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, Marketing Science Institute, Cambridge, Mass.*

Engel, James F., and Blackwell, Roger D. 1982. Consumer Behavior Hinsdale, Ill.: Dryden Press.

Helson, Harry. 1959. "Adaptation Level Theory," in Sigmund Koch (ed.), *Psychology: A Study of Science*, 1, New York: McGraw-Hill. Howard, John A., and Sheth, Jagdish N. 1969. *The Theory of Buyer Behavior*, New York: John Wiley.

Juster, Thomas F. 1966. "Consumer Buying Intentions and Purchase Probabilities; An Experiment in Survey Design, "*Journal of the American Statistical Association*, 61 (September), 658-96.

Kim, Jae-An and Kahout, Frank J. 1975. "Special Topics in General Linear Models," in Nie et. al., (Ed.), *Statistical Package for the Social Sciences*, New York: McGraw-Hill.

Lingoes, James C., and Pfaff, Martin. 1972. "The Index of Consumer Satisfaction Methodology," Proceedings, Association for Consumer Research, 689-712.

Lord, Frederick M. 1963. "Elementary Models for Measuring Change," in C.W. Harris (Ed.), *Problems in Measuring Change*, Madison: University of Wisconsin Press, 22-38.

Magnusson, David. 1965, *Test Theory*, Palo Alto, Calif.: Addison Wesley. Oliver, Richard L. 1980. "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions", *Journal of Marketing Research*, Vol. XVII (November) pp. 460-469.

Thibaut, J.W., and Kelley, H.H. 1959. *The Social Psychology of Groups*, New York: John Wiley.

ABOUT THE AUTHOR

VED PRAKASH is Assistant Professor of Marketing, Florida International University. He received his D.B.A. degree in Marketing from the University of Tennessee, Knoxville. The areas of his teaching and research interests are Consumer Satisfaction, Product Familiarity, Involvement, Mood States, Psychological Gender Differences and Advertising Format, and Negative Information. His work has appeared in the *Journal of the Academy of Marketing Science*, in a book on *Personal Values, Advances in Consumer Research*, Proceedings of National *AIDS*, and Proceedings of the *Academy of Marketing Science*. Most recently, he has completed a study of the relationship between Values, Expectations and Consumer Satisfaction.