A REASSESSMENT OF THE RELIABILITY OF DIFFERENCE SCORES IN THE MEASUREMENT OF DISCONFIRMATION OF EXPECTATIONS

Richard A. Spreng, Michigan State University

ABSTRACT

Difference scores have been utilized in past tests of the disconfirmation of expectations model, and are currently used in measurement scales such as SERVQUAL. However, difference scores have been criticized as having low reliability. For example, Prakash and Lounsbury (1983) found very low reliabilities for disconfirmation scores for two products. It is argued here that there are several problems with their analysis that indicate that the reliability of difference scores may not be as low as they claim. An empirical example demonstrates reasonably good reliabilities for disconfirmation measures based on difference scores.

INTRODUCTION

The concept of disconfirmation of expectations has dominated research on consumer satisfaction. Generally, disconfirmation has been measured in one of two ways: subtractive disconfirmation and subjective disconfirmation. Early research calculated the difference between a consumer's pre-use expectations of the product and the consumer's post-use perception of product performance. Recently, researchers have generally used subjective measures of disconfirmation, in which consumers are asked if the product performed better or worse than they expected (Bearden and Teel 1983; Churchill and Surprenant 1982; Oliver 1980; Tse and Wilton 1988). There are several arguments made for following this latter approach. First, it is suggested that disconfirmation "represents a distinct psychological construct encompassing a subjective evaluation of the difference between product performance and the comparison standard..." (Tse and Wilton 1988, p. 205). A second reason advanced for using subjective disconfirmation is that when a model includes direct effects of expectations, disconfirmation, and performance on satisfaction, and disconfirmation is measured as the difference between expectations and performance, multicolinearity is likely to be a problem. This is avoided by measuring disconfirmation independently.

However, subjective disconfirmation measures have been criticized for several reasons (Cadotte, Woodruff and Jenkins 1987). First, by not measuring expectations and perceptions of performance, managers are limited in explaining why disconfirmation occurred (e.g., were expectations too high or was the performance too low). Cadotte et al. also criticized many measures used in satisfaction since they often contain evaluative anchors (e.g., good/bad) that imply a standard, and can result in the confounding of measures of expectations, perceived performance, and disconfirmation. Finally, subtractive disconfirmation appears to be closer to the conceptual definition of disconfirmation.

A number of problems with subtractive disconfirmation measures have also been identified. Specifically, several authors have criticized difference scores in general (Johns 1981; Peter, Churchill and Brown 1993). For example, Johns (1981) suggested that difference scores have a number of problems, including lack of theory relating the two components, the difference score often does not add anything "more" than the component scores, and the low reliability of difference scores. The issue of the advisability of using difference scores rather than a subjective assessment of the difference will not be dealt with in this paper. Nevertheless, it should be recognized that there are situations in which the use of difference scores may be valuable. For example, when researchers are using certain standards (e.g., consumer desires), the level of the standard (e.g., what consumers desire) may be of interest. If one measures the component scores (desires and performance) *and* a subjective measure of the difference, each attribute will need to be measured three times. Utilizing a difference score would necessitate only two sets of measures. Thus, in situations in which measures of the component scores are needed, operationalizing the discrepancy between a standard and actual

performance may be more efficiently done with a difference score. This type of information may be of greater importance in early stages of a customer satisfaction measurement program.

One of the problems most often associated with difference score measures is that they are purported to have low reliability. The more general problem of low reliability of difference scores is important, since a recent set of measurement scales for service quality (Zeithaml, Parasuraman and Berry 1990) uses difference scores. The purpose of this paper is to reexamine the reliability issue of difference scores.

RELIABILITY OF DIFFERENCE SCORES

The reliability problem can be easily seen by examining the formula for the reliability of a difference score (Lord 1963):

$$r_{diff} = \frac{\sigma^2_1 r_{11} + \sigma^2_2 r_{22} - 2r_{12}\sigma_1\sigma_2}{\sigma^2_1 + \sigma^2_2 - 2r_{12}\sigma_1\sigma_2}$$

Where:

r_{diff} = reliability of the difference score
σ^2_1 = variance of the first component
σ^2_2 = variance of the second component
r_{11} = reliability of first component
r_{22} = reliability of the second component
r_{12} = correlation between the two components

It is clear from this formula that reliability will decrease as the reliability of either of the components goes down, or as the correlation between the components increases. Prakash and Lounsbury (1983) provided an empirical example in which measures of expectations were combined with measures of perceptions of performance, and the reliability of the resulting difference score was low for two products (.46 for a restaurant, and .19 for beer).

The Correlation Between the Components

Based on the above formula, one way to increase the reliability would be to decrease the correlation between the component measures. In fact, in the example provided by Prakash and Lounsbury (1983), the correlation between expectations and performance was .66 for the restaurant and .61 for the beer. Further, with regard to trying to decrease the correlation between the components, they state "..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." However, what is being measured is a construct, such as expectations, on an attribute or set of attributes. If expectations and perceptions of performance are indeed distinct constructs, there is no reason to assume that they need to be highly correlated, just because the same attribute is being measured.

Johns (1981) discusses this point in terms of the situation in which the same construct is measured from independent sources, such as managers and subordinates. Johns suggests that it would be reasonable to expect these measures to be correlated, in that "... a lack of correlation between the component parts of difference scores would call into question the (convergent) validity of the components themselves, raising doubts as to whether they refer to the same stimulus object." But Johns was referring to the same construct (e.g., manager's reports of how frequently their superiors expected them to perform various job activities, and superiors' reports of how frequently they in fact wanted the job activities performed) as perceived by different people. Thus, both measures deal with superiors' desires regarding job activities, with one measure being managers' perception of these desires, and the other being supervisors' actual desires. This is quite a different situation than that involved in difference scores of expectations and perceptions of product performance, in that here we are dealing with different constructs.

There are, however, several reasons that expectations and perceptions of performance may be correlated. The first is the well established effect of expectations on performance (Olshavsky and Miller 1972), in that perceptions of performance move in the direction of expectations, particularly with respect to hard to judge outcomes. A second reason deals with the use of measurement scales that share an evaluative aspect, as criticized by Cadotte et al. (1987). Satisfaction

research often utilizes expectations measures that are evaluative in nature, and perceptions of performance generally have a very evaluative component.

The Reliabilities of the Components

A second reason the reliabilities in the Prakash and Lounsbury example were low was due to moderate reliabilities of the components. For the restaurant, the reliabilities were .83 and .81 for expectations and performance respectively. The reliabilities for the beer product were .65 for expectations and .71 for performance.

Prakash and Lounsbury used 11 attributes for the restaurant, and seven attributes for the beer product. Rather than calculate a difference score for each attribute, they summed the expectations measures over all 11 attributes, summed the performance measures over all attributes, and calculated the difference between these two summations. This was necessary to be able to calculate the component reliabilities by Cronbach's coefficient alpha. However, since coefficient alpha assumes that the measures are multiple items of the same construct, summing the items is valid only if the attributes are all highly correlated. The attributes they use (e.g., food served hot, menu variety, location convenience) are not multiple measures of a single construct, but single measures of different constructs (i.e., expectations about how hot the food will be, expectations about how varied the menu will be, etc.). By summing very different attributes that may not be highly related, and calculating coefficient alpha for these measures, the reliability may appear to be rather low. Nunnally (1978, p. 212) states that "... reliability depends entirely on the average correlation among items and the number of items." The empirical example presented next will illustrate this. The reliability estimates for difference scores for three items that represent different attributes will be compared with the reliability estimates for difference scores for three items that represent a single attribute.

EMPIRICAL EXAMPLE

Overview

The following data are from a larger study of consumer satisfaction. The subjects were 211 adult volunteers from a church in a midwestern community. The subjects ranged in age from 18 to over 65, with a median age in the 31 to 35 age group. The product used was a camcorder.

Method

Subjects received an expectations manipulation, and completed pre-use questions, including measures of expectations, on five attributes: ease of use, versatility, picture quality, picture sharpness, and picture colors. Subjects were then exposed to the performance manipulation, in which they used the camcorder. Some subjects used a camcorder that was completely automatic, and saw a test tape that indicated that the performance was not very good in difficult video situations (e.g., when a backlight control is needed). Other subjects used the same brand of camcorder, but the manual overrides had to be used, making the camcorder difficult to use. These subjects saw a test tape that showed that the camcorder performed well in several difficult video situations. Manipulation checks indicated that both the expectations and performance manipulations succeeded. Both expectations and performance were measured as beliefs (e.g., "What level of performance do you anticipate or expect from this camcorder," anchored by "Not very easy to use/Very easy to use").

Reliability Analysis

To illustrate the points made above, the reliability of difference scores for three attributes that are dissimilar will be compared with the reliability of difference scores based on three similar attributes.

Ease of use, versatility, and picture quality were chosen as three attributes that are fairly different, and picture quality, picture sharpness, and picture colors were chosen to represent multiple measures of overall picture outcome. The Table shows the coefficient alphas for expectations

and perceptions of performance, the correlations between the constructs, and the reliabilities for the difference scores.

Table 1
Coefficient Alpha, Correlations and Reliabilities

Alpha Expectations	Alpha Performance	Correlation between Exp. and Perf.	Reliability of Difference Score
Dissimilar Items			
.88	.80	.34	.77
Similar Items			
.99	.97	.36	.97

There are several differences between the present results and those from Prakash and Lounsbury. First, in the present study the correlations between the constructs, while significant, are much lower than those in Prakash and Lounsbury. This accounts for the moderate reliability with even the dissimilar set of items. Second, the coefficient alpha, not surprisingly, is much higher when the items used share a similar aspect of the product. The end result is that difference scores based on this more homogeneous set of items is very high, even though the number of items is quite small.

DISCUSSION

Prakash and Lounsbury (1983) seemed to recognize the point made in this paper, in that they recommend one way to increase the reliability of difference scores is to increase the homogeneity of the items in the component measures. Johns (1981) also suggests that combining items measuring different attributes into an index is problematic. But even more importantly, difference scores should be calculated on an attribute basis, rather than as differences of sums.

Second, if distinct constructs are being combined in a difference score, then there is no reason to worry about a low correlation between the constructs. On the contrary, measures should be constructed, as suggested by Cadotte, et al. (1987), so that they are not confounded with each other.

A number of approaches to measuring customer satisfaction involve the comparison between the consumer's perception of the product performance and a standard, such as expectations (Oliver 1980), desires (Olshavsky and Spreng 1989), product norms (Cadotte, et al. 1987), "an excellent company" (Zeithaml et al. 1990) and others. This paper demonstrates that difference scores may not have the low reliability that has been suggested in the past.

REFERENCES

Bearden, W. O. and J. E. Teel (1983), "Some Determinants of Consumer Satisfaction and Complaint Reports," *Journal of Marketing Research*, 20, (February), 21-28.

Cadotte, Ernest R., Robert B. Woodruff and Roger L. Jenkins (1987), "Expectations and Norms in Models of Consumer Satisfaction," *Journal of Marketing Research*, 24, (August), 305-14.

Churchill, G. A. and C. Surprenant (1982), "An Investigation into the Determinants of Consumer Satisfaction," *Journal of Marketing Research*, 19, (November), 491-504.

Johns, Gary (1981), "Difference Score Measures of Organizational Behavior Variables: A Critique," *Organization Behavior and Human Performance*, 27, 443-463.

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.

Nunnally, Jum C. (1978), *Psychometric Theory*, New York: McGraw-Hill.

Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17 (November), 460-469.

Olshavsky, Richard W. and John A. Miller (1972), "Consumer Expectations, Product Performance, and Perceived Product Quality," *Journal of Marketing Research*, 9, (February), 19-21.

Olshavsky, Richard W. and Richard A. Spreng (1989), "A 'Desires as Standard' Model of Consumer Satisfaction," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 2, 49-54.

Peter, J. Paul, Gilbert A. Churchill, Jr. and Tom J. Brown (1993), "Caution in the Use of Difference Scores in Consumer Research," *Journal of Consumer Research*, 19, (March), 655-662.

Prakash, Ved and John W. Lounsbury (1983), "A Reliability Problem in the Measurement of Disconfirmation of Expectations," *Advances in Consumer Research*, 10, 244-249.

Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25, (May), 204-12.

Zeithaml, Valarie A., A. Parasuraman and Leonard L. Berry (1990), *Delivering Quality Service*, New York: Free Press.

Send correspondence regarding this article to:
Richard A. Spreng
Department of Marketing & Logistics
Michigan State University
311 Eppley Center
East Lansing, MI 48824