

Seymour Sudman University of Illinois
Survey Research Laboratory, 909 W. Oregon, Suite 300, Urbana, IL 61801

Keywords: Satisfaction, Questionnaire Evaluation, Surveys

As these three interesting papers indicate, measuring satisfaction of customers and users has become not only a very important substantive issue, but has raised important methodological concerns. In my discussion, I will attempt to emphasize and perhaps extend a bit the consideration of some selected methodological points.

I'm going to discuss the papers in somewhat different order than they were presented, to reflect the timing of the various stages of the process, starting with the mainly conceptual paper by Sharma and his colleagues, turning next to the Fowler paper that deals with the development of a satisfaction questionnaire, and ending with the paper by Devlin and her colleagues that deals primarily with the testing and validation of satisfaction questionnaires. I was intrigued by the fact that the papers although really very different have some shared similarities.

The papers by Sharma and his colleagues and Devlin and her colleagues both make it clear that among the first steps necessary in developing measures of consumer satisfaction one must identify whom to ask the questions and what to ask them about. Although the framework proposed by Sharma et al. is a general one, the interesting example given refers to a situation where the customer is not an individual, but an organization. In this situation, very careful attention must be given to who are the key personnel in the customer organization who will be contacted. Obviously, this will vary from organization to organization, but the important point made is that almost always using multiple informants will be superior to using a single informant

The second significant issue addressed in the Sharma et al. paper, as well as all the other papers is what are the important process drivers of satisfaction. In the paper by Sharma and his colleagues, these process drivers were initially identified by management and management judgment was confirmed in later multivariate analysis. While this method worked fine in the example given, it is better, in my judgment, to do as Fowler and Devlin et al. suggest, that is, initial focus groups or depth interviews with customers to make sure that you are not missing any key points. On the other hand, the multivariate methods employed are powerful indicators of the relative importance of alternative drivers of satisfaction, and can clearly tell a user company where it should focus its efforts.

All three papers raise the issue as to whether satisfac-

tion should be based on a specific transaction or on some global cumulative measure, and come down on different sides of the issue reflecting the same split in satisfaction studies. I think Sharma and his colleagues make a good case for overall satisfaction when interactions between a firm and a customer are frequent and irregular so that monitoring individual events is impractical. But note that the outcome of using the global satisfaction measures is to point to possible changes in organizational structure and not to evaluate the effectiveness of individuals or groups.

I have an unfair advantage in discussing the Fowler paper, since I was involved in critiquing earlier versions of the questionnaire he discussed. Rather than discussing the individual questions and the decisions that led to them, I'd like to step back and consider in somewhat greater detail than Jack did two of the methodological issues he raised that are also raised in the other papers.

1. Should events or overall satisfaction be measured?
2. What to ask about?

Jack did not make it clear in his presentation that this questionnaire is intended to be used primarily as a mail survey that is sent to respondents' homes. One can certainly imagine other ways of obtaining satisfaction data. Thus, one alternative would be to gather the data at the site where the care was provided. The data could be obtained either by an interviewer or by a self-administered questionnaire. Current technology would allow for either a computer assisted questionnaire where respondents would read answers from the screen or an audio computer assisted questionnaire where the recorded questions would be read to them over a set of earphones.

Certainly, if the sampling is on-site, the most natural questions to ask are about the visit that has just occurred, although more global satisfaction questions are, of course, possible. As we heard in the Devlin et al. paper, this strategy of discussing a specific event is the one used in the large, continuous satisfaction studies done by the phone companies. On the other hand, global satisfaction measures are used by J.D. Powers in their studies of satisfaction with automobiles and by the researchers at the University of Michigan who are studying multiple product categories. When are specific measures more useful? The more specific measures, I believe, are better suited for evaluating and changing the behavior of managers and other key personnel within an organization. The focus is on micro issues that may or may not need adjustment. The more global measures are better for modeling customer loyalty and future switching behavior, and for changing

organizational structures as pointed out by Sharma et al.

I am much more comfortable with the analysis of trend data of global measures than in making strong statements about what the levels mean. Jack said that the aim of the Consumer Assessment of Health Plans was to compare across plans so that people making choices could benefit from the experiences of those already enrolled, but this aim makes the strong assumption that any observed differences are real, and not the result of sample differences. Differences in many sample variables such as geography, age, ethnic background and current health status can certainly impact on global rating, let alone differences in response rates at different sites, especially if different modes are used. The use of trends, as in the paper by Sharma and his colleagues, to measure the effectiveness of an organizational change, makes good sense.

This blends in with the second issue-what should be asked? Here, I believe the approach presented by Jack is very much the right one-the use of focus groups and cognitive interviews to determine what are important and meaningful questions to ask, rather than having the provider attempt to construct them unaided. Note that it isn't simply asking questions about issues that are important to the user. It is also necessary that the user be able to make an informed judgment. Thus, while competence of the physician or the repairman may be of utmost importance, it is often difficult or impossible for the patient or user to determine this, until perhaps at a much later date.

Susan and her colleagues suggest a range of methods for testing questionnaires and scales, ranging from simple observations to the use of psychometric tools. It is always useful to be reminded that observing the data carefully is a necessary first step before turning to more powerful statistical methods. Thus, they sensibly point out that too many "don't know" or "scale not used" responses would suggest the question may not be meaningful to respondents, and that bimodal or unexpected response patterns may indicate confusion or difficulty with the concept being tested. They warn about positive response bias limiting the power of the scale to discriminate, but this will frequently be a problem in satisfaction surveys-the great majority of respondents will be very satisfied with the service they received and only a few will have any complaints. Efforts to improve the discriminating power of scales may sometimes simply add noise or measurement error without increasing the reliability.

They make an important point in suggesting the value of scale responses can be tested by

1. Evaluating the tone of open-end responses to a "why do you feel that way?" question.

2. Assessing the consistency between the scale response and the report of what occurred, i.e., the actual length of time the respondent had to wait for a repairman or doctor and the scale rating on promptness scale.

I must admit to being less taken with Susan's and her colleagues' suggestions on how to detect respondent irritation and confusion. The proposed methods-interviewing interviewers, looking for excessive dropout rates, tracking repeated response patterns and bizarre verbatims could all be effective for a really terrible question or questionnaire, but are less likely to be very helpful for moderately flawed instruments that have some problems, but have been prepared by researchers with some experience. Thus, hardly ever would be a questionnaire be so terrible that more than a handful of respondents would drop out in the middle, and the bizarre verbatims which so lighten our lives when we see or hear them are very rare. Here, the Fowler approach of focus groups and cognitive interviews seems to me much more direct.

Susan is on solid psychometric ground and in agreement with Subhash when she proposes the use of factor analysis to identify key drivers of satisfaction, and the use of Chronbach's alpha as a measure of internal consistency of a scale. As she points out, this is a method for getting rid of questions that are unrelated to the factors that determine satisfaction.

As she also points out, satisfaction surveys virtually never measure the same customer twice for reliability because it is costly, time consuming, and inconvenient. The thought occurred to me that one way of doing this would be in situations where the questionnaire is event based such as a physician visit or repair call, and where the respondent reports exactly what occurred as well as a satisfaction response. For those respondents who had two events within a specified period and reported the same thing occurred both times, i.e., a half hour wait for the doctor or a two-day wait for the repairman, the correlation in scale responses would be an indicator of reliability.

Susan and her colleagues specifically raise the issue of non-response bias that I mentioned earlier in Jack's paper. Assuming that the non-response bias remains relatively constant over time, tracking trends in results rather than relying on the data from one point in time is the best thing to do. If the biases change over time, because of a change in methods, survey organization or for some other reason, then all bets are off. It is impossible to parcel out real changes in satisfaction from sample or method changes unless the old and new series have been deliberately overlapped so that chain-linking is possible.

To sum up, these are three solid papers and they all have important things to tell us about measuring customer satisfaction.