

Problematic Scales

When measuring quality, expectations scales exhibit several drawbacks.

By Terry Grapentine

Measuring customer "expectations" has gained some popularity in building models to assess product quality and customer satisfaction. However, interpretation of findings from surveys that use "expectations" and/or "requirements" scales demonstrate how their use can be problematic, at times to the point of being ludicrous. When measuring product quality, alternative scaling methods -- particularly those based on performance -- may be more appropriate.

The issues of customer satisfaction and product and service quality have received considerable attention from the marketing community in recent years, sparking many articles and books, as well as numerous conferences and seminars on these topics.

In a 1993 *Marketing Research* article entitled "Selecting a Scale for Measuring Quality," Susan J. Devlin, H.K. Dong, and Marbue Brown proposed the use of two "expectations" scales in measuring the perceived quality of products and services: a 5-point expectations scale and a 4-point requirements scale (see Exhibit 1). Although they claim that "the perfect rating scale doesn't exist," with regard to the aforementioned two scales they said:

- "Both scales work well in telephone and paper delivery."
- "Both scales have been well-received in companies where they have been introduced because they link measures to current definitions and philosophy about total quality management."

QUESTIONABLE VALIDITY

Although much of the Devlin et al. discussion represents a potentially useful resource for the marketing research practitioner, the validity of these two particular scales is questionable because of the problems associated with the following:

- Tracking perceptual changes over time.
- Handling ideal point attributes.
- Handling vector attributes.
- Using multivariate statistical analysis.
- Scale truncation.
- Conceptual definitions of "expectations" and "requirements."

Tracking Perceptual Changes Over Time

Exhibit 2 gives measurements for a hypothetical respondent between two time periods using both the expectations and requirements scales. In both instances, the scale values increased by one point (which is a negative finding).

This certainly would be a concern for management because it suggests that, in the case of the expectations scale, product performance declined from "better than expected = 2" to "just as expected = 3." Similarly, for the requirements scale, product performance declined from "met requirements = 2" to "nearly met requirements = 3." These data are not actionable, however, because interpretation of the data is ambiguous.

For example, with respect to the expectations scale, what could account for a change in ratings? Did perceived product performance remain the same and customer expectations increase? Did product performance decline and customer expectations stay constant? Or, did product performance decline and customer expectations increase?

A similar problem is associated with the requirements scale. What can account for a change in this scale's ratings between time period T_1 and T_2 ? Did perceived product performance remain the same and customer requirements increase? Did perceived product performance decline and consumers' requirements stay constant? Or, did perceived product performance decline and consumers' requirements increase?

In short, the expectations and requirements scales may provide ambiguous results in tracking studies.

Handling Ideal Point Attributes

Two general categories of product attributes sometimes are referred to as ideal point or vector attributes. An ideal point attribute is one in which there is an "ideal amount" or "optimal" amount of an attribute. Receiving either less or more than the ideal amount of the attribute reduces perceptions of quality. A vector attribute, on the other hand, is one in which more of the attribute is better.

There are many examples of ideal point attributes in the field of marketing research. For example:

Exhibit 1

<i>Devlin et al. expectations scales</i>	
5-point expectations scale	4-point requirements scale
1 = Much better	1 = Exceeded
2 = Better	2 = Met
3 = Just as [expected]	3 = Nearly met
4 = Worse	4 = Missed
5 = Much worse	

- In the marketing of bank services, there is an ideal frequency with which a bank account officer should call on a client. Too few or too many calls can affect the customer's perceptions of service quality.
- In distribution studies, ideal point attributes are often used to measure the frequency with which raw materials are delivered to a manufacturer. Because of the nature of the manufacturing process, it is possible for shipments to be received "too soon" or "too late."
- In retail marketing, a store can be too unattractive (suggesting the store carries low-quality merchandise) or too attractive (suggesting the store is too high priced).

Both the expectations and requirements scales fail to capture the relationship between ideal point attributes and perceived quality because of how respondents conceptually answer these scales. For respondents to use the requirements scale, they need to perform three tasks. First, a respondent must determine the required level of the attribute in

Exhibit 2

<i>Tracking perceptual changes over time</i>		
Time period of collecting survey data	Scale measurements for one respondent	
	Expectations scale	Requirements scale
T_1	2	2
T_2	3	3
Difference $T_2 - T_1$	1	1

Exhibit 3

Handling ideal point attributes					
Perceived (P) "level" of attribute X		(P - R)			
P	4	3	2	1	0
	3	2	1	0	-1
	2	1	0	-1	-2
	1	0	-1	-2	-3
Hypothetical ideal level *		1	2	3	4
		Required (R) "level" of attribute X			

*The assumed ideal level = 2.

question (attribute X). Second, he or she must evaluate the perceived level of attribute X. And third, the respondent must make a mental subtraction between the required and perceived levels of attribute X delivered in order to articulate whether the supplier exceeded, met, nearly met, or missed the respondent's requirements.

Exhibit 3 shows potential problems associated with using the requirements scale when ideal point attributes are involved. Different attribute "levels" have been noted by the scale values 1-4, where higher numbers reflect more of the attribute. (The 4-point scale in this example was chosen arbitrarily; any multiresponse scale would serve here.)

Assume we are conducting a service quality study for a bank. The client wants to know whether the attractiveness of the bank's facilities exceeded, met, nearly met, or missed the customers' requirements for quality. In this hypothetical example, the ideal level of attribute X is denoted by the number 2 because the customer does not want the bank to spend too much money on making the bank's facilities attractive. However, if the bank is perceived performing at level 3 or 4, the respondent is forced to answer that the bank "exceeded" his or her requirements.

This would be a non-optimal situation because the attribute in question is an ideal point attribute. Yet, the requirements scale suggests that the bank exceeded the customer's requirements and that such a situation would be positively correlated with perceived quality when, in fact, just the opposite is true. (For a detailed discussion of a similar problem associated with

another expectation scale -- SERVQUAL -- see articles by R. Kenneth Teas published in 1993 and 1994 in the *Journal of Marketing*.)

Handling Vector Attributes

Exhibit 4 identifies a problem associated with using the expectations scale when measuring vector attributes. Recall that, for a respondent to answer this expectation scale, he or she must make a mental calculation of the difference between the perceived level (P) of product performance and the expected level (E) of product performance.

The exhibit shows four different scenarios in which the customers' expectations would be met (that is, $P - E = 0$.) In effect, one could have four very different respondents indicate that their expectations were "met" under four different situations:

- $P = 1; E = 1; (P - E) = 0$.
- $P = 2; E = 2; (P - E) = 0$.
- $P = 3; E = 3; (P - E) = 0$.
- $P = 4; E = 4; (P - E) = 0$.

Devlin et al. indicate that such a situation severely compromises the usefulness of a scale: "If diverse service experiences translate to one response category, information is lost."

Exhibit 4 shows this indeed to be the case with both the expectations and requirements scales because four different service experiences translate to one response category. Such a finding suggests that the following two situations produce equal "quality":

- $P = 1; E = 1; (P - E) = 0$
- $P = 4; E = 4; (P - E) = 0$.

These two situations cannot produce the same level of perceived quality because: (1) Vector attributes assume that higher numbers on the scale denote higher levels (i.e., better) performance, and (2) a situation in which $P = 4$ is, by definition, a higher quality of performance than when $P = 1$.

An implicit assumption underlying the expectations and requirements scales is that a company can increase perceived quality by keeping

product performance the same but reducing consumers' expectations or requirements. If a marketer's goal is to increase perceived product quality, why would he or she want to diminish customers' expectations or requirements as opposed to developing strategies designed to enhance the perceived quality?

Statistical Analysis

Presumably, the reason one would use the expectations or requirements scales is to identify those attributes which are most predictive of product or service quality. A standard research method used for this purpose is regression analysis.

According to Devlin et al., "The using of multivariate statistical methods, such as factor analysis, logistic regression, and covariant structure analysis, helps assess which scale delivers the highest reliability and validity measures, reduces multicollinearity concerns, and has the greatest prediction power of criterion measures (e.g., overall quality or loyalty)."

In Exhibit 4, however, a fundamental assumption underlying the use of regression analysis has been violated because diverse service experiences can be translated to one response category. Consequently, predictor variables using the expectations and/or requirements scale have serious measurement validity problems.

Scale Truncation

Scale truncation reflects a situation in which it is impossible for a respondent to use certain portions of a scale. For example, when we briefly tested the requirements scale, some respondents found it very difficult to conceptualize a situation in which a bank could "exceed" a customer's requirements for quality on attributes such as (1) "The bank's monthly statements are accurate" and (2) "The bank has banking hours that are convenient for my family."

Regarding the first attribute, some respondents indicated that degrees of "accurateness" do not exist, only degrees of inaccuracy. As one respondent said: "It would be impossible for a bank to exceed my

Exhibit 4

Problems associated with measuring vector attributes

Perceived (P) "level" of attribute X		(P - E)			
	4	3	2	1	0
P	3	2	1	0	-1
	2	1	0	-1	-2
	1	0	-1	-2	-3
		1	2	3	4
		Expected (E) "level" of attribute X			

requirements of accuracy. The statements are either accurate or they are not."

A similar situation was found with regard to the second attribute. Some respondents found it difficult to use the scale because exceeding their requirements would provide no extra utility. As one respondent said: "It's irrelevant for a bank to exceed my requirements on this statement. As long as the bank is open when I need it [to be] open, that's all I care about. If it's open at 3:00 in the morning, so what?"

Depending upon the attributes used in the survey, it might be impossible for a company (or product) to "exceed" the customer's requirements for quality, thus truncating the scale.

Conceptual Definitions

Perhaps the most serious problem associated with using these two scales is the fact that the terms "expectations" and "requirements" are ambiguous to respondents and can introduce significant measurement error in the data.

Expectations: What does the term expectations mean? Is an "expectation" a measurement of forecasted performance or is it a measurement of the minimum acceptable performance level of an attribute?

In using the expectation scale, a respondent could indicate that the performance on an attribute was "much better" than expected. But this rating could take on the following interpretations:

- The performance is "much better" because the forecasted performance exceeded historical performance.
- Current performance is "much better" because actual performance exceeded the minimal performance level required by the customer.

Potential problems with regard to the ambiguity of the expectations concept have received considerable attention in the recent marketing research literature.

Requirements: Similar problems are associated with measuring the requirements concept. Assume we are asking respondents whether a certain supplier's performance on attribute X exceeded, met, or missed their requirements for quality.

In this situation, researchers have found that the term "requirements" can mean: "What customers feel they deserve," "minimal acceptable performance," or "ideal performance."

In summary, the terms expectations and requirements are ambiguous and have not been defined clearly by Devlin et al., who purport to use these terms in measuring perceived quality.

PERFORMANCE SUPERIOR

Offering an alternative measurement scale to the Devlin et al. scales without discussing a totally different conceptual framework for measuring quality is difficult. This is because their scales assume that the "expectation" and/or "requirement" concepts are important variables in predicting quality, and such constructs have severe validity problems.

The Teas' model of evaluated performance (EP), when compared to models that incorporate expectations measurements, is superior in predicting perceived quality. Briefly, the EP model implies that the perceived quality of a product or service can be increased by closing the gap between the product's/service's performance and that of the ideal product/service on an attribute.

Thus, a scale that measures perceived product or service *performance* may be more predictive of

overall quality than a scale that measures expectations.

A scale that could be used to measure perceived performance is one in which respondents are given a list of attributes and asked to use a scale from 0-10 to indicate how well the attribute *describes* the product/service. A rating of 10 denotes the statement very much describes the product/service, and a rating of 0 indicates the statement does not at all describe the product/service.

Note that a similar measurement framework (one that measures only perceived product/service performance and not "expectations") was used in Teas' 1993 *JM* article, which compared pairwise correlations between several quality models and global measures of product quality and satisfaction. Quality models that incorporated measures of perceived product performance had higher validity coefficients than models that also incorporated measures of expectations.

The 0-10 point "describes"/"does not describe" scale, however, is not effective when measuring ideal point attributes. One method that can be used to measure an ideal point attribute is to have respondents evaluate the performance of a service, for example, on a unit of measurement that reflects different "amounts" of the service. This type of scale is called an intensity scale.

For example, in measuring the ideal frequency with which a corporate account calling officer should make a personal visit to a customer's place of business, one could ask respondents the ideal frequency with which these visits should occur.

Then, respondents could be asked the actual frequency with which these visits do occur. One could then subtract the actual from the ideal frequency to examine the "gap" between actual and ideal performance.

An alternative approach is to use a scale that measures evaluated performance. For example, if a performance attribute is described on an evaluative continuum -- such as poor performance vs. excellent performance -- the assumption can be made that infinite ideal points are involved. That is, excellent

performance can be assumed to be preferred over poor performance.

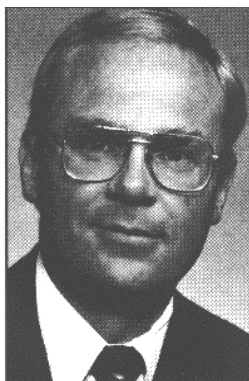
This is in contrast to intensity-type scales in which the amount of the attribute is measured. Once performance is measured on an evaluative scale, the ideal point is assumed to be infinite, eliminating the problem and therefore the need to measure ideal points.

Perceived performance scales, evaluated performance scales, or intensity scales might provide

better measures of quality than do expectations or requirements scales. However, the Devlin et al. process of testing the reliability and validity of scales is an important contribution to the literature and can be used in future theoretical and applied research to discover how scales in general should be constructed and administered to develop the most valid measures of quality. [MR]

ADDITIONAL READING

- Churchill, Gilbert A. Jr. and J. Paul Peter (1984), "Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis," *Journal of Marketing*, 21 (November), 360-75.
- Cronin, Joseph J. Jr. and Michael H. Morris (1989), "Satisfying Customer Expectations: The Effect on Conflict and Repurchase Intentions in Industrial Marketing Channels," *Journal of the Academy of Marketing Science*, 17 (Winter), 41-9.
- _____ and Steven A. Taylor (1992), "Measuring Service Quality: A Re-examination and Extension," *Journal of Marketing*, 56 (July), 55-68.
- Devlin, Susan J., H.K. Dong, and Marbue Brown (1993), "Selecting a Scale for Measuring Quality," *Marketing Research: A Magazine of Management & Applications*, 5 (Summer), 12-17.
- Liechty, M. and Gilbert A. Churchill Jr. (1979), "Conceptual Insights into Consumer Satisfaction With Services," in *Educators' Conference Proceedings*, Series 94, Neil Beckwith et al., eds. Chicago: American Marketing Association, 509-15.
- Miller, John A. (1977), "Studying Satisfaction, Modifying Models, Eliciting Expectations, Posing Problems, and Making Meaningful Measurements," in *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, H. Keith Hunt, ed. Cambridge, MA., Marketing Science Institute, 72-91.
- Parasuraman, A., Valerie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (Spring), 12-40.
- Teas, R. Kenneth (1993), "Expectations, Performance Evaluation and Consumers' Perceptions of Quality," *Journal of Marketing*, 57 (October), 18-34.
- _____ (1993), "Consumer Expectations and the Measurement of Perceived Service Quality," *Journal of Professional Services Marketing*, 8 (2), 33-54.
- _____ (1994), "Expectations as Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment," *Journal of Marketing*, 58 (January) 132-139.
- Tse, David K. and Peter C. Wilton (1988), "Models of Consumer Satisfaction Formation: An Extension," *Journal of Marketing Research*, 25 (May), 204-212.
- Wittink, Dick R., and Leonard R. Bayer (1994), "Statistical Analysis of Customer Satisfaction Data: Results from a National Experiment with Measurement Scales," working paper #94-04, Cornell University Johnson Graduate School of Management.



Terry Grapentine is Founder and Principal of Grapentine Co. Inc., a marketing research firm based in Ankeny, Iowa