ROBERT A. WESTBROOK

A simple yet effective rating scale measure of satisfaction has been developed by sociological researchers studying the perceived quality of life. The author examines the suitability of this measure for marketing studies of consumer satisfaction/dissatisfaction. Reliability, validity, and selected measurement properties are evaluated empirically.

A RATING SCALE FOR MEASURING PRODUCT/ SERVICE SATISFACTION

RECENT sociological research on the perceived quality of life (Andrews and Withey 1976) suggests a promising new measure for the study of consumer satisfaction—the Delighted-Terrible (D-T) scale. The measure has been used successfully in assessing evaluative responses to major life concerns such as one's job, health, family, and neighborhood. As applied to consumer products and services, the D-T scale has several potential advantages in comparison with the rating scale methods typically employed to study consumer satisfaction. This article examines the suitability of the D-T scale for consumer satisfaction applications. Progress in the measurement of consumer satisfaction would benefit not only applied studies in industry and government, but also more basic research on the process through which consumers arrive at such judgments about their purchases.

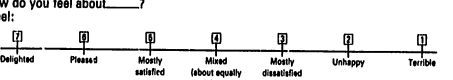
Despite the complexity of the construct of satisfaction, consumer researchers have used rather simple measures, most often single-item rating scales of four to seven points between the extremes of "very satisfied" and "very dissatisfied" (e.g., Andreasen and Best 1977; Ash 1978; Handy 1977). The fact that these measures commonly yield very skewed distributions of responses, suggests that the scales may be insufficiently sensitive to detect gradations of consumers' sentiments. In addition, there is some evidence that such rating scales indicate higher levels of satisfaction than analyses of the content of free responses to unstructured questions (Andreasen 1977). Unfortunately, few validation studies for simple, commonly used "satisfied-dissatisfied" rating scales have been attempted, and comprehensive measure comparisons as reported by Haley and Case (1979) for attitude scales are sorely needed.

The potential advantages of the D-T scale (illustrated in Figure 1) include (1) improved representation of the construct of satisfaction through more explicit reference to and gradation of the affective component, (2) improved differentiation of responses at the upper end of the satisfaction continuum, and (3) allowance for respondents who may never have evaluated their satisfaction with the product/service, which reduces potential response bias due to obtrusiveness and demand effects (Day 1977). Despite these advantages, the D-T scale has yet to gain widespread application in consumer research. Though the scale was validated for Andrews and Withey's sociological research, its suitability for the study of consumer satisfaction is not known. This article examines its reliability, validity, and selected measurement properties.

Robert A. Westbrook is Associate Professor of Marketing, University of Arizona.

Journal of Marketing
Vol. 44 (Fall 1980), 68-72.

FIGURE 1
Alternative Measures of Product/Service Satisfaction

Measure	Description
D-T Scale	How do you feel about_____? I feel:  A Neutral (neither satisfied nor dissatisfied) B I never thought about it
Percentage Scale	Overall, how satisfied have you been with this_____? 100% Completely Satisfied — 90 — 80 — 70 — 60 — 50 — 40 — 30 — 20 — 10 — 0% Not at all Satisfied
Need S-D	To what extent does this _____ meet your needs at this time? Extremely Well (7) : _____ : _____ : _____ : _____ : _____ : _____ (1) Extremely Poorly
Content Analytic	Coding of free responses to a series of unstructured questions* into the following categories: 1. Only unfavorable evaluations 2. Both favorable and unfavorable evaluations 3. Neither favorable nor unfavorable evaluations 4. Only favorable evaluations

*The questioning procedure consisted of a general, nondirective question on subjects' thoughts and feelings about the experiences of owning and using the product, followed by clarifying probes, and finally a specific open-ended followup question about any particular aspects of the product liked or disliked.

Method

Sources of Data

Three separate studies were conducted to obtain data from different consumer populations over a variety of products and services. In Study I, the focus was automobiles, banks, and wristwatches. Self-administered questionnaires were completed by 72 upper-level undergraduate and graduate students of business at a major southwestern state university. Respondents were questioned only about those products actually owned and used. Study II consisted of personal interviews with 151 female heads of household in two large southeastern cities. Though judgmental area sampling procedures were used, sample demographic characteristics were similar to those of area residents according to the U.S. Census. Eighty-seven respondents were questioned about their washing machines, 32 about their refrigerators, and 32 about their color televisions. In Study III, self-administered questionnaires were completed by 47 undergraduates at the same state university. Respondents were questioned only about their automobiles to allow gathering of other data relevant to measure validation.

Validation Procedures

In assessing the suitability of the D-T scale for product/service satisfaction applications, the following issues were of major concern.

1. *Reliability.* To what extent does the measure yield consistent estimates? Reliability was estimated by the test-retest method over a 10-day interval. Despite its limitations, it is the only means of assessing reliability for single-item measures.
2. *Convergent Validity.* To what extent does the measure correlate with dissimilar measures of satisfaction for the same item? The alternative satisfaction measures examined in this research are shown in Figure 1.
3. *Discriminant Validity.* To what extent is the measure unique, or discriminable from measures of other constructs? Suitable discriminability is indicated by low correlations between the measure

and (a) equivalent measures of unrelated constructs, and (b) dissimilar measures of theoretically unrelated constructs. Equivalent measures of unrelated constructs were obtained by using the D-T scale to assess satisfaction with other products owned by subjects. As the research did not include any truly dissimilar measures of satisfaction (e.g., a physiological measure), the investigation of discriminant validity is based solely on criterion (a) Evidence with regard to criterion (b) must await the development of truly dissimilar measures of satisfaction.

4. *Nomological Validity.* To what extent does the measure behave as it should, according to theoretical predictions? To address this issue, various well-accepted antecedents to and effects of satisfaction were selected from theory. Nomological validity is indicated if the measure supports hypothesized relationships between satisfaction and its antecedents and effects. Hypotheses and their respective operational definitions are shown in Figure 2.

Findings

Test-retest estimates of reliability for the D-T scale are shown in the diagonal of the multitrait-multimethod matrix in Table 1. They range from .65 (automobiles) to 84 (banking services). Though modest, these figures are sufficiently large to encourage further examination of validity.

FIGURE 2
Hypotheses About Antecedents and Effects of Satisfaction

Hypothesis	Variable	Theoretical Relation to Satisfaction	Measure
H_1	Realization of positively-valued expectations	Direct	Graphic rating scale: 1 = Not as good as expected 2 3 = About as expected 4 5 = Much better than expected
H_2	Realization of negatively-valued expectations	Inverse	Graphic rating scale: 1 = Not as serious as expected 2 3 = About as expected 4 5 = Much more serious than expected
H_3	Extent of product repairs	Inverse	Total reported expenditures for parts and labor to repair product
H_4	Possession of desired product features	Direct	"Does your (*product*) have all the features and conveniences you'd like, or would you prefer it had certain things it doesn't now have?" IF YES: "Which ones?"
H_5	Composition of evoked set	Direct	"When you next buy another (*product*), are there any particular brands you will *definitely* consider buying?" IF YES: "Which ones?"
H_6	Composition of inert set	Inverse	"When you next buy another (*product*), are there any brands you will *definitely not* consider buying?" IF YES: "Which ones?"
H_7	Complaint activity	Inverse	An index constructed by assigning one point for mention of each of the following: (a) complaints to retail outlet (b) complaints to manufacturer (c) complaints to government agency (d) complaints to friends, relatives

TABLE 1
Multitrait-Multimethod Matrix: Correlations Between Rating Scale Measures of Satisfaction (Study I)

Satisfaction Measure	Product	D-T Scale			% Scale			Need S-D		
		Auto	Bank	Watch	Auto	Bank	Watch	Auto	Bank	Watch
D-T Scale	Auto	(.65)[a]								
	Bank	−.19	(.84)[a]							
	Watch	−.08	−.01	(.73)[a]						
% Scale	Auto	.85[a]	−.19	.14	(.55)[a]					
	Bank	−.04	.81[a]	−.04	−.05	(.81)[a]				
	Watch	−.02	.19	.65[a]	−.10	.11	(.72)[a]			
Need S-D	Auto	.48[a]	−.11	.15	.42[a]	.07	.06	(.68)[a]		
	Bank	−.14	.84[a]	−.06	−.15	.87[a]	.15	.02	(.78)[a]	
	Watch	−.02	−.06	.81[a]	−.07	−.16	.57[a]	.16	−.07	(.69)[a]

Note: Parenthesized values refer to reliability estimates. Bases for correlations range from 58 to 70 because of differences in product/service ownership by respondents and item nonresponse.
[a]Significant at .01 level.

TABLE 2
Correlations Between D-T Satisfaction Ratings and Hypothesized Antecedents and Effects

Hypothesis	Variable	Study I			Study II		
		Automobiles (N=63)	Banks (N=72)	Watches (N=60)	Washing Machines (N=87)	Refrigerators (N=32)	Color Televisions (N=32)
H_1	Realization of expectations for favorable outcomes	.23	.57	.39			
H_2	Realization of expectations for unfavorable outcomes	.25	.57	.46			
H_3	Frequency of repairs				−.24	−.49	−.51
H_4	Possession of desired features				.27	.21	.28
H_5	Evoked set composition				.38	.50	.50
H_6	Inert set composition				−.48	−.47	−.62
H_7	Complaint activity				−.44	−.26	−.68

Note: All values significant at .05 level except H_4 for refrigerators and color televisions. Empty cells indicate data not available because of study design.

Convergent validity of the D-T scale is indicated by the large and significant correlations in Table 1 between alternative satisfaction measures within product categories. Studies II and III yield similarly high correlations between the D-T scale and the percentage rating scale, ranging from .65 to .86 depending on product. In addition, Study II indicates convergence between the latter two rating scales and a content analytic measure based on subjects' free response to a series of unstructured questions (see Figure 1). Correlations between the D-T scale and the content analytic measure range from .73 to .78 depending on product (all were major appliances); those between the percentage rating scale and the content analytic measure range from .54 to .86.

One of the conditions for discriminant validity requires that satisfaction with a given product, as measured by the D-T scale, not correlate with satisfaction ratings for other products also obtained by this measure. The correlations in Table 1 indicate that D-T ratings of satisfaction with automobiles, banks, and wristwatches

all are independent. Similarly, in Study II, D-T ratings of satisfaction across the set of major appliances owned by respondents are uncorrelated; detailed data are not shown because of space limitations.

Evidence of nomological validity is shown by the relationships of the D-T scale to its hypothesized antecedents and effects (Table 2). Study I data support both H_1 and H_2 for all products. In Study II, H_3 through H_7, all are confirmed for washing machines, and all but H_4 for refrigerators and color televisions. Even in the latter cases, however, observed correlations are in the predicted direction.

Though mean satisfaction ratings are typically high across product categories (ranging from 4.82 to 5.80), analysis of the distribution of responses reveals an appreciable reduction in the proportion lying at the upper extreme of the satisfaction continuum in comparison with the percentage scale (which is similar to the rating scales typically used in previous research). The skewness statistics[1] follow.

	D-T Scale	*Percentage Scale*
Automobiles (Study I)	−.155	−.171
Automobiles (Study II)	− .63	−1.90
Washing machines	−1.20	−2.14
Refrigerators	−1.19	−1.24
Color TV	−1.18	−2.06
Wristwatches	−1.07	−2.20
Banks	− .70	− .90

To appraise the significance of these differences, a one-tailed *t*-test for matched observations was performed. The differences are significant at the $\alpha = .02$ level ($t = 3.18$, 6 d.f.). These results may be attributed to the superiority of the D-T scale in discriminating the affective component of satisfaction. For example, of 25 washing machine owners in Study II who selected the uppermost scale position on the percentage scale ("completely [10]%] satisfied"), only 14 chose the uppermost scale position on the D-T scale ("delighted"), the remaining 11 respondents describing their satisfaction as "pleased" or "mostly satisfied." Similar results were obtained for other categories.

Conclusion

Evidence from three separate studies with durable goods and banking service indicates the suitability of the D-T scale for measurement of consumer satisfaction. The measure has reasonable reliability, converges with other rating scales and free-response measures. In addition, it behaves as would be expected of a measure of satisfaction. Finally, the D-T scale reduces the skewness of satisfaction responses typically noted in previous studies.[2] Overall, these findings should not only encourage further use of the D-T scale in both basic and applied studies, but also allay concerns about the quality of satisfaction measurement (Engledow 1977).

Despite the predominance of simple, single-item measures of satisfaction, future research should focus on the development of multi-item measures. They could reduce measurement error—enhancing reliability—and permit improved representation of the construct of interest. Given the complexity of the concept of satisfaction, such developments should be accorded high priority.

[1]Skewness represents the third moment about the mean of a distribution. For details see Hays (1973, p. 248).

[2]Though problems of skewness can be handled analytically by transformations after data have been collected (e.g., see Winer 1972, p. 397), it may be preferable to avert these complications by the use of more appropriate measures.

REFERENCES

Andreasen, Alan R. (1977), "A Taxonomy of Consumer Satisfaction/Dissatisfaction Measures," *Journal of Consumer Affairs*, II, no. 2 (Winter), 11–24.

——— and Arthur Best (1977), "Consumers Complain—Does Business Listen?", *Harvard Business Review* (July–August), 93–101.

Andrews, Frank M. and Stephen B. Withey (1976), *Social Indicators of Well-Being*, New York: Plenum Press.

Ash, Stephen B. (1978), "A Comprehensive Study of Consumer Satisfaction with Durable Products," in *Advances in Consumer Research*, 5, H. K. Hunt, ed., Ann Arbor: Michigan: Association for Consumer Research, 254–62.

Day, Ralph L. (1977), "Extending the Concept of Consumer Satisfaction," in *Advances in Consumer Research*, 4, W. D. Perreault, ed., Atlanta: Association for Consumer Research, 149–54.

——— ed. (1979), *Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, Bloomington, Indiana: Indiana University Press.

Engledow, Jack (1977), "Was Consumer Satisfaction a Pig in a Poke?", *Business Horizons*, 20 (April), 87–94.

Haley, Russell I. and Peter B. Case (1979), "Testing Thirteen Attitude Scales for Agreement and Brand Discrimination," *Journal of Marketing*, 43 (Fall), 20–32.

Handy, Charles R. (1977) "Monitoring Consumer Satisfaction with Food Products," in H. K. Hunt, ed., *Conceptualization and Measurement Satisfaction and Dissatisfaction*, Cambridge, Mass.: Marketing Science Institute.

Hays, W. L. (1973), *Statistics for the Social Sciences*, 2nd ed., New York: Holt, Rinehart and Winston.

Winer, B. J. (1972), *Statistical Principles in Experimental Design*, New York: McGraw-Hill Book Company.