Marketing Letters 8:2 (1997): 239–250

Measuring the Importance of Positive Constructs: A Test of Alternative Rating Procedures

JOHN A. MCCARTY
Kogod College of Business Administration, American University, 4400 Massachusetts Avenue, N.W., Washington, D.C. 20016
e-mail: mccarty@american.edu

L. J. SHRUM
Department of Marketing, Rutgers University, 220 Levin Bldg., Rockafeller Rd., New Brunswick, NJ 08903
e-mail: shrum@everest.rutgers.edu

Abstract

When constructs are positive in nature, individuals rating a set of them in terms of importance tend to end-pile their ratings toward the positive end of the scale, resulting in little differentiation among the items rated. Results of two experiments indicate that a procedure where respondents first pick their most and least important items in a set and then rate the entire set of items (*least-most*) and a procedure where respondents rank the items prior to rating (*rank-then-rate*) provide more differentiation and less end-piling than a simple rating procedure (*rate-only*). The results also show that the increased differentiation for the least-most method results in improved fit of latent structure, compared to a rate-only procedure. These results generalize across two types of personal values scales (Rokeach Value Survey and Kahle's List of Values scale), number of items rated (9 and 18) and number of rating points (1 to 10 and 0 to 100).

Key words: values, measurement, differentiation, scales

Introduction

The rating of a construct along a continuum of importance is central to marketing research. However, problems may arise when the items being rated are all viewed as inherently positive by raters. In such cases, even though the rating scale provided contains typical positive and negative anchors (such as extremely important/not at all important), the raters will tend to end-pile their ratings toward the positive end of the scale and thus show little differentiation among the rated items. This lack of differentiation can affect statistical relations, both among the items themselves (such as affecting factor analyses and measurement) and between the items and other variables (such as affecting predictive ability).

This article presents the results of two experiments that are intended to address problems encountered when rating positive constructs. Using personal values as the positive construct, two alternative rating techniques were tested for their ability to reduce end-piling and increase differentiation.

Personal values and marketing

In the last two decades, researchers have increasingly turned attention toward the relation of personal values to consumer behavior (see Homer and Kahle, 1988, for a review). Furthermore, researchers in marketing and consumer behavior have suggested that personal values may be a useful market segmentation tool (Kamakura and Mazzon, 1991; Kamakura and Novak, 1992). Such a position is consistent with trends in industry, as evidenced by the popularity of the VALS instrument (Holman, 1984). The primary premise is that people differ on fundamentally held personal values, and these stable beliefs underlie a variety of consumption behaviors; therefore, these differences are potentially useful as a means of segmenting and targeting potential consumers.

Studies relating personal values to consumer behavior have generally shown modest but consistent relationships between values and various aspects of behavior. Although it may be that the lack of strong findings are simply the result of weak relationships between values and consumer variables, it may just as well be that the weak findings are attributable to problems with the measurement of values. Because personal values are relatively abstract concepts, the proper measurement of them has proven to be a difficult task, as evidenced by the number of studies that have addressed measurement issues. Although some studies have focused on competing value instruments (e.g., Beatty et al., 1985; Kahle, Beatty, and Homer, 1986; Novak and MacEvoy, 1990), the most prevalent measurement issue has focused on whether respondents should *rank* the values from most to least important or *rate* each value on an interval scale of importance.

Ranking and rating methods each has its own set of advantages and disadvantages (for a more thorough discussion of the pros and cons of each method, see Alwin and Krosnick, 1985; Krosnick and Alwin, 1988). Rankings force choice between values, whereas ratings allow for ties. Rankings tend to be more time-consuming and difficult to administer (for example, with telephone interviews) and provide only ordinal level data, but ratings are relatively easy to administer and provide interval level data.[1] However, one of the bigger drawbacks to using ratings for values measurement, and one that is the focus of the studies to be presented here, concerns the lack of differentiation that respondents make among the values. A rating procedure does not require respondents to make choices among the values in a set. Because of the inherent positive nature of values, few people consider any of the values to be *un*important. Consequently, respondents may tend to end-pile their ratings toward the positive end of the scale, with little differentiation among the value items, making it difficult to ascertain the relative importance of the values to the respondent.

The issue of lack of differentiation with rating data is important because it has the potential to affect statistical relations among variables. For example, correlations between values and behavior depend in part on the variability of the values measurements. Consequently, low correlations may not necessarily be due to weak relationships but instead may result simply because the measurements of the values did not vary to a sufficient extent to detect their influence (Wyer, Bodenhausen, and Gorman, 1985). Similarly, lack of variability in the value measurements may potentially affect correlations among the values themselves, which in turn may affect factor analyses and causal modeling results.

Despite the large number of studies that have addressed the ranking versus rating debate, previous research has been equivocal regarding which method is the better approach to measuring personal values (Krosnick and Alwin, 1988). In this article we are interested in developing and evaluating rating methods that encourage respondents to compare values and make distinctions among them similar to that of a ranking method, yet provide the interval level data of rating procedures. A review of recent values studies makes it clear that consumer researchers, as well as values researchers in other fields, are increasingly adopting a rating procedure as the method of choice, regardless of the rank versus rate debate (Shrum and McCarty, in press). Moreover, because the lack of differentiation in the rating of personal values is likely due to their inherent positive nature (rather than an aspect of the construct itself), such alternative rating methods should also be useful for rating tasks involving other constructs including importance ratings of products, services, and brands, as well as consumer satisfaction ratings. Thus, the development of alternative rating methods can be beneficial to researchers in the area of values as well as other domains of consumer behavior.

Alternative rating procedures

A desirable property of a ranking method is that it requires respondents to make choices among the values, thus making salient the relative importance of the values to the respondent. It seems likely that rating methods could be improved if respondents are encouraged to make their ratings within a context of choice. If the relative importance of the values is made salient to respondents, it is likely that they would discriminate more among the values when they rate them.

There are two particular rating procedures that may provide such a choice context and thus help circumvent the problems of end-piling and lack of differentiation. A procedure that has respondents first rank the values before rating them (rank-then-rate) has been suggested (Munson, 1984) and used (e.g., Crosby, Bitner, and Gill, 1990; Schwartz and Bilsky, 1990; Shrum, McCarty, and Loeffler, 1990). This procedure requires respondents to make choices and comparisons among values through the ranking task. With these choices and comparisons still fresh in memory, the respondents then rate the values.

Although a rank-then-rate procedure would seem intuitively preferable to a rate-only method in that it forces comparisons but also overcomes the problems with ordinal measurements, the procedure obviously suffers from other problems associated with rankings—namely, time and effort. This is particularly troublesome because the lengthy ranking part of the task is done merely to force respondents to consider all of the values prior to the rating task; ratings are used for the analysis.

A desirable procedure would be one in which respondents are forced to consider all of the values prior to rating but one that does not require the exceptional amount of time and energy of a rank-then-rate task. A "least-most" rating procedure is one possible approach. This procedure, similar to an anchoring technique considered by Munson and McIntyre (1979), asks respondents to first look through the entire set of values and choose the most important value. Next, the respondents are asked to again look through the values and

choose the least important value. Such a task forces the respondent to observe and scan the entire list and make comparisons among the values. After respondents select their most and least important values from the list, they are then requested to rate the entire set of values on an interval scale.

The procedure takes only a little more time than a rate-only method but requires much less time than the lengthy process of ranking the values and then rating them.[2] Like the rank-then-rate task, least-most requires respondents to read the entire list and make comparisons in order to determine the least important and most important value. The question then becomes whether the least-most procedure induces respondents to make comparisons to a sufficient extent that it results in differentiation similar to rank-then-rate, and whether the least-most procedure does in fact induce more differentiation than the rate-only method.

Overview of the experiments

Two experiments were conducted to assess whether the alternative rating procedures (rank-then-rate and least-most) are preferable to a rate-only procedure in terms of end-piling and differentiation. Experiment 1 compared all three methods (rank-then-rate, least-most, and rate-only) and used the Rokeach Value Survey (Rokeach, 1973). Experiment 2 compared least-most to rate-only, and used Kahle's List of Values scale (Kahle, 1983).

For both experiments, degree of end-piling was operationalized as the average of the importance ratings across all values for each respondent and the percent of the values rated with the highest rating point (percent top box). Greater end-piling is associated with higher average importance ratings and a greater percent of values receiving the highest rating point. Differentiation was conceptualized as the degree to which respondents make distinctions among the values. Two measures that each tap a somewhat different aspect of differentiation were computed: (1) the standard deviation of the importance ratings for each respondent, and (2) P_d (rho), which represents the likelihood that a respondent will distinguish among the values (Krosnick and Alwin, 1988; Linville, Salovey, and Fischer, 1986). The formula for P_d is given by the following equation:

$$P_d = 1 - \sum_{i=1,n} P_i^2$$

where P_i represents the percent of the values rated at a given point on a rating scale; i varies from the lowest point on a particular rating scale to the highest possible rating point. P_d is maximized when a respondent assigns a different rating to each of the values and is minimized when a respondent assigns the same rating to all values. Thus, P_d varies as a function of the number of different scale points that a respondent uses. The standard deviation, on the other hand, varies as a function of the distance between the scale points used by the respondent and is maximized if a respondent rates half of the values with the highest rating point and the other half with the lowest rating point.

If the alternative rating procedures do indeed result in less end-piling and greater differentiation than a simple rating of the values, then we would expect the following: the alternative rating procedures will result in less end-piling than rate-only, indicated by a lower average importance rating across all values and a lower percentage of top box scores; the alternative rating procedures will result in a greater range of scale points used compared to rate-only, indicated by a larger average standard deviation of the importance ratings; and the alternative rating procedures will result in more scale points used, compared to rate-only, indicated by a larger average P_d.

We were also interested in whether differences in differentiation would affect statistical relations among the values that were rated. If in fact the least-most procedure produces greater differentiation among the values, we would expect that the assessment of latent structure of the values would result in a better model fit for the least-most procedure than the rate-only method. This would suggest that the least-most method provides better quality data than that of the rate-only procedure. Experiment 2 evaluated the fit of the latent structures of the LOV value scale provided by the least-most and the rate-only method.

Experiment 1

Method

Sample. The data for Experiment 1 were collected from undergraduate students at a large university in either a classroom situation or during an experimental hour. The total sample consisted of 197 respondents, with 62 in the rate-only condition, 73 in the rank-then-rate condition, and 62 in the least-most condition.

Measures and procedure. The terminal set of the Rokeach Value Survey (RVS) was used as the value measurement scale. The terminal scale consists of eighteen values, which Rokeach (1973) describes as measuring desirable end-states of existence (such as happiness, equality, and an exciting life).

The difference in the written instructions for the rating task represented the experimental manipulation of the study. In the rank-then-rate condition, respondents were directed to first rank the set of values in terms of importance, then rate the importance of each value on a 0 to 100 scale. In the least-most condition, respondents were instructed to scan the list of values and select the one that was most important to them, then scan the list and select the one that was the least important. They were then instructed to rate the importance of the values on the 0 to 100 scale. In the rate-only approach, respondents were simply asked to rate each of the eighteen values on the 0 to 100 scale.

Results and discussion

Means for the importance ratings, percent of top box, standard deviation, and P_d were calculated across respondents in each of the measurement conditions. An analysis of

variance was performed on each of the measures to determine group differences for the three rating conditions; because the comparsions between groups were planned, Least Significant Difference multiple comparsion tests were performed to determine group differences.

The top portion of Table 1 shows the means for each of the dependent variables for Experiment 1. Degree of end-piling, indicated by the average importance rating, differed as a function of rating method, ($p < .0001$). As expected, those in the rate-only condition rated the values higher than those in the least-most condition ($p < .001$) and the rank-then-rate condition ($p < .001$), indicating that both the least-most and rate-only rating methods reduced end-piling. However, the means of the least-most and rank-then-rate procedures were not significantly different ($p > .05$), suggesting that these methods resulted in similar degrees of end-piling. Similar results were obtained when end-piling was operationalized as the percentage of values rated with the highest rating point ($p < .025$). Those in the rank-then-rate and least-most conditions did not differ ($p > .05$), but both were significantly lower than rate-only ($p < .05$).

The rating methods also differed in terms of the amount of differentiation they induced. When differentiation was operationalized as the within respondent standard deviation, the overall analysis of variance showed significant group differences as hypothesized,

Table 1. End-piling and differentiation as a function of rating method for Experiments 1 and 2.

Experiment 1				
Experimental Condition	Average Importance Rating	Percent of Values Rated 100 (Top Box)	Average Rating Standard Deviation	Average Rho
Rank-then-rate	75.42[a]	12.0%[a]	19.43[a]	.86[a]
	(10.24)	(13.9%)	(7.64)	(.07)
Least-most	78.49[a]	13.0%[a]	17.91[a]	.84[a,b]
	(8.93)	(14.2%)	(7.16)	(.08)
Rate only	84.87	19.1%	12.87	.81[b]
	(8.77)	(17.5%)	(6.66)	(.09)
Percent of difference between rate-only and rank-then-rate yielded by least-most	63.5%	85.9%	76.8%	60.0%
Experiment 2				
Experimental Condition	Average Importance Rating	Percent of Values Rated 10 (Top Box)	Average Rating Standard Deviation	Average Rho
Least-most	8.08	27.2%	1.48	.67
	(1.12)	(22.8%)	(.60)	(.15)
Rate only	8.72	38.6%	1.00	.56
	(.83)	(30.6%)	(.52)	(.20)

Note: Standard deviations are in parentheses. Means sharing same superscript are not significantly different from each other.

($p < .0001$). Those in the rate-only condition exhibited less differentiation than those in the least-most condition ($p < .001$) and the rank-then-rate condition ($p < .001$), suggesting that the rate-only method results in less dispersion of ratings than least-most and rank-then-rate. Also, as expected, no difference in differentiation was observed between the least-most and rank-then-rate conditions ($p > .05$). Similarly, when differentiation was measured by P_d, the overall analysis of variance indicated significant differences in P_d as a function of rating method ($p < .002$). P_d was significantly greater for the rank-then-rate group than the rate-only group ($p < .001$), indicating that respondents in the rank-then-rate condition tended to use more rating points than those in the rate-only group. However, the mean P_d for those in the least-most condition fell between the other two means and was not significantly different from either group ($p > .05$).

This pattern of findings suggests that both least-most and rank-then-rate tend to induce more differentiation and less end-piling than rate-only, although there was no significant difference in P_d between rate-only and least-most. Additionally, for the most part, rank-then-rate and least-most performed similarly in inducing differentiation. Indeed, as Table 1 shows, the least-most rating method yielded 60 percent or more of the improvement in discrimination (decrease in end-piling; increase in differentiation) provided by the rank-then-rate method relative to the rate-only method for all four of the dependent measures.

Experiment 2

Although the findings of Experiment 1 indicate that level of differentiation varies as a function of rating procedure, it may be that the results are peculiar to the particular values instrument used (the RVS) or the large number of scale points used in the rating instrument (101 points). Experiment 2 was designed to extend the previous findings to a different values measurement instrument having fewer values to be rated and fewer scale points. We were also interested in whether the differences in differentiation due to the use of different rating methods produce changes in the latent structures of the values.

Method

Sample. The sample consisted of 394 students from a large university. The data were collected during an experimental hour that partially fulfilled a course requirement. Of the total sample, 195 students were in the least-most condition and 199 were in the rate-only condition.

Measures and procedure. Respondents completed the LOV scale, which is comprised of nine Rokeach-like terminal values. This scale has been suggested to be quicker and easier to complete and to be more relevant to consumer behavior issues than the larger Rokeach scale (Beatty et al., 1985). Each respondent completed either the rate-only or least-most task under instructions similar to those of Experiment 1. However, in this experiment, the rating scale required judgments to be made on a 1 to 10 scale from "very unimportant" to

"very important." The average importance rating, percent of top box, standard deviation, and P_d were again used as the dependent measures.

Results and discussion

End-piling and differentiation. The means for each of the dependent variables in Experiment 2 are shown in the bottom of Table 1. As in Experiment 1, respondents in the rate-only condition tended to end-pile their ratings more than those in the least-most condition, both in terms of average importance, ($p < .001$) and percent of values using the highest rating point, ($p < .001$). In terms of differentiation, the least-most condition produced a higher average standard deviation of the importance ratings (more dispersion) than respondents in the rate-only condition, ($p < .001$), and those in the least-most condition also exhibited a higher P_d (used more scale points) than those in the rate-only condition, ($p < .001$).

These results are consistent with Experiment 1 and suggest that the least-most method induces more differentiation and less end-piling in the ratings of the values. Furthermore, the results indicate that this effect is not merely an artifact of a particular scale, the number of values rated, or the number of scale points used.

Latent structure. Simultaneous factor analysis using the generalized least-squares estimation method via EQS (Bentler, 1993) was employed to evaluate the two rating methods. The recommended procedure to test the similarities of factor structures begins with a test of the equality of the covariance matrices of the two groups (Jóreskog, 1971). If the matrices are found to be the same, no further tests are necessary as the separate covariance matrices would be essentially the same as the pooled covariance matrix. If the hypothesis of equality of the two covariance matrices is found to be untenable, however, a series of tests of the similarity of factor structures are made. The first test determines whether the two groups have the same number of factors with the same pattern of loadings. If so, then more stringent tests are made. Specifically, after the test of the same number of factors with the same pattern, the hypotheses proceed as follows: (1) the test of an invariant factor pattern matrix, (2) the test of an invariant factor pattern and the same unique variances, and (3) the test of an invariant factor pattern, the same unique variances, and an invariant factor variance-covariance matrix. Because these hypotheses are ordered by level of stringency, if one test is found to be untenable, the tests that would follow are unnecessary.

We also analyzed the factor structure resulting from each rating method separately in order to show any differences in quality of fit. The measures of goodness of fit we used were χ^2, the Satorra-Bentler scaled χ^2 (Bentler, 1993), which is robust to deviations from multivariate normality, the centrality index (McDonald and Marsh, 1990), and the nonnormed fit index (Bentler and Bonett, 1980). Because values data are typically not multivariate-normally-distributed and χ^2 is affected by sample size, our evaluation of fit focuses on the Satorra-Bentler scaled χ^2, the centrality index, and the nonnormed fit index.

The data from both of the measurement methods were evaluated in terms of the extent to which they fit a three-factor model for the LOV scale. The hypothesized factors were those that had been empirically determined in previous studies (see Grunert, Grunert, and Beatty, 1989; Kennedy, Best, and Kahle, 1989). The values of self-fulfillment, self-respect, and a sense of accomplishment were expected to comprise one factor; excitement, fun and enjoyment in life, and warm relationships with others were expected to load on a second factor; a sense of belonging, being well-respected, and security were expected to load on a third factor.

The hypothesis of the equality of covariance matrices was tested and found to be untenable, $\chi^2 = 174.72$, $df = 45$, $p < .001$. We then proceeded to test whether the data from the rate-only group and the least-most group had latent structures with the same pattern of loadings on the same number of common factors, using the factor model just described. The factor loadings for each of the rating methods are presented in Table 2. The chi-square statistic for the two groups considered simultaneously indicates that the model did not fit both groups well ($\chi^2 = 105.52$, $df = 48$, $p < .001$). The scaled chi-square,

Table 2. Standardized parameter estimates for three-factor model of the LOV scale for Experiment 2.

	Parameter Estimates for Least-Most Group *Factor*			Parameter Estimates for Rate-Only Group *Factor*		
	1	2	3	1	2	3
Factor 1						
Self-fulfillment	.690			.634		
Self-respect	.913			.723		
Sense of accomplishment	.781			.749		
Factor 2:						
Excitement		.548			.701	
Warm relationships		.755			.477	
Fun and enjoyment		.736			.736	
Factor 3:						
Sense of belonging			.393			.402
Being well-respected			.592			.775
Security			.595			.746
Factor correlations:	1	2	3	1	2	3
1	1.000			1.000		
2	.750	1.000		.670	1.000	
3	.759	.667	1.000	.736	.521	1.000
Goodness of fit for both groups considered simultaneously:	$\chi^2 = 105.52$, $df = 48$, $p < .001$					
Goodness of fit for each group:						
χ^2	38.984, $df = 24$, $p = .027$			66.900, $df = 24$, $p < .001$		
Scaled χ^2	37.622, $df = 24$, $p = .038$			45.747, $df = 24$, $p < .005$		
Centrality index	.962			.897		
Nonnormed fit index	.993			.977		

Note: All loadings are significant at $p < .05$. Factor loadings fixed at zero during analysis are omitted from table.

centrality index, and nonnormed fit index all suggest a strong fit of the model for the least-most rating procedure. However, although the factor loadings for the rate-only group were significant and the nonnormed fit index was high, the scaled chi-square and centrality index suggest that the data from the rate-only procedure did not fit the hypothesized model as well as the least-most rating procedure. These results suggest that the increase in differentiation that results from the least-most procedure results in a better fit of the measurement model.

General discussion

Both the rank-then-rate and least-most methods force respondents to read through all of the values and make comparisons and contrasts among the values in the set. In contrast, the rate-only method allows respondents to avoid such comparisons. The results of the two experiments presented here indicate that the least-most and rank-then-rate procedures increase the differentiation of value ratings compared to a rate-only method, both in terms of dispersion (as indicated by the rating standard deviation) and the number of different rating points used (as indicated by P_d). This finding indicates that having respondents compare and contrast the values, either by first ranking them, or simply by having them select their most and least important values, increases the psychological distance respondents perceive among the values.

The results of the comparison of latent structures between least-most and rate-only suggest that rating method may affect more than simply the degree to which respondents distinguish among the values. The findings of differences in quality of fit in latent structure between the two rating procedures from Experiments 2 indicate that the least-most procedure generally provides better measurement.

These findings are potentially important for researchers who are interested in measuring personal values and determining their relation to other consumer variables. If typical rate-only methods of measuring values are employed, it is possible that important relations between values and other variables of interest may be missed due to the restriction of range induced by a rate-only method. On the other hand, methods such as rank-then-rate and least-most may increase the detection of relationships between values and criterion variables of interest. The results of Experiment 2 suggest that increased differentiation will provide for a better fit when analyzing latent structure, and this increase in fit should also affect the relations between value factors and criterion variables.

However, we do not want to overstate our case. The differences we observed were not always large. It is possible that other techniques that build on this research may be developed to provide even greater differentiation than the methods we tested. On the other hand, it is certainly plausible that relations between values and criterion variables may be strengthened at least to some degree by the use of a least-most or rank-then-rate method.

The advantages of the least-most approach as a rating procedure were demonstrated with two currently available value scales and for different numbers of rating points. These results suggest that the procedure represents a general technique that will be useful for a variety of value rating instruments and may also prove useful in the development of new

instruments. Moreover, this rating method may be helpful in the rating of other types of stimuli (such as brand attributes and consumer satisfaction) where lack of differentiation is a problem because of the positive nature of the stimuli.

Future research should focus on at least two issues. First, the results should be replicated on a more generalizable sample. The two studies here used only student samples. Second, studies should test whether the increases in differentiation produced by the alternative rating methods do in fact result in stronger or more revealing relations with criterion variables. Ideally, such criterion variables should include those that have been used in previous studies (such as demographics), so that comparisons can be made to some previous standard. These types of comparisons should provide the ultimate test of the usefulness of the alternative rating methods evaluated here.

Notes

1. Although we use the term *interval-level data*, we recognize that from a psychometric perspective, rating data may not be precisely interval-level. However, such data (for example, from Likert and semantic differential scales) are typically treated as interval level for data analysis and most social science studies refer to such data as interval-level.
2. The relative time required to complete the Rokeach Value Survey for the three different rating methods (rank-then-rate, least-most, and rate-only) was tested with seventy-eight respondents. Respondents completing the RVS by the rank-then-rate method took an average of 626.25 seconds, those completing it via the least-most method required an average of 397.56 seconds, and those using the rate-only method averaged 300.34 seconds to complete the survey (all of these mean times were significant from each other, $p < .05$). The rank-then-rate method thus requires more than twice as much time to complete as the rate-only method, whereas the least-most technique requires only about one-third more time to complete as the rate-only method.

References

Alwin, Duane F., and Jon A. Krosnick. (1985). "The Measurement of Values in Surveys: A Comparison of Ratings and Rankings." *Public Opinion Quarterly* 49, 535–552.

Beatty, Sharon E., Lynn R. Kahle, Pamela Homer, and Shekhar Misra. (1985). "Alternative Measurement Approaches to Consumer Values: The List of Values and the Rokeach Value Survey." *Psychology and Marketing* 2, 181–200.

Bentler, Peter M. (1993). *EQS Structural Equations Program Manual*. Los Angeles: BMDP Statistical Software.

Bentler, Peter M., and Douglas G. Bonett. (1980). "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88, 588–606.

Crosby, Lawrence A., Mary Jo Bitner, and James D. Gill. (1990). "Organizational Structure of Values." *Journal of Business Research* 20, 123–134.

Grunert, Suzanne C., Klaus G. Grunert, and Sharon E. Beatty. (1989). "Measuring Personal Values in Consumer Research: The Cross-Cultural Validity of the List of Values." In Tadeusz Tyxzka and Piotr Gasparski (eds.), *Homo Oeconomicus: Presumptions and Facts—Proceedings of the Fourteenth Annual Colloquium of the International Association for Research in Economic Psychology* (vol. 17) (pp. 471–487). Kazimierz Dolny, Poland: Polish Academy of Sciences.

Holman, Rebecca. (1984). "A Values and Life Styles Perspective on Human Behavior." In Robert E. Pitts, Jr. and Arch G. Woodside (eds.), *Personal Values and Consumer Psychology* (pp. 35–54). Lexington, MA: Lexington Books.

Homer, Pamela, and Lynn R. Kahle. (1988). "A Structural Equation Test of the Values-Attitude-Behavior Hierarchy." *Journal of Personality and Social Psychology* 54, 638–646.

Jöreskog, Karl G. (1971). "Simultaneous Factor Analysis in Several Populations." *Psychometrika* 36, 409–426.

Kahle, Lynn R. (ed.). (1983). *Social Values and Social Change: Adaptation to Life in America*. New York: Praeger.

Kahle, Lynn R., Sharon E. Beatty, and Pamela Homer. (1986). "Alternative Measurement Approaches to Consumer Values: The List of Values (LOV) and Values and Life Style (VALS)." *Journal of Consumer Research* 13, 405–409.

Kamakura, Wagner A., and Jose Afonso Mazzon. (1991). "Value Segmentation: A Model for the Measurement of Values and Value Systems." *Journal of Consumer Research* 18, 208–218.

Kamakura, Wagner A., and Thomas P. Novak. (1992). "Value-System Segmentation: Exploring the Meaning of LOV." *Journal of Consumer Research* 19, 119–132.

Kennedy, Patricia, Roger J. Best, and Lynn R. Kahle. (1989). "An Alternative Method of Measuring Value-Based Segmentation and Advertising Positioning." In James H. Leigh and Claude R. Martin (eds.), *Current Issues and Research on Advertising* (pp. 139–155). Ann Arbor: Division of Research, University of Michigan Business School.

Krosnick, Jon A., and Duane F. Alwin. (1988). "A Test of the Form-Resistant Correlation Hypothesis." *Public Opinion Quarterly* 52, 526–538.

Linville, Patricia W., Peter Salovey, and Gregory W. Fischer. (1986). "Stereotyping and Perceived Distributions of Social Characteristics: An Application to Ingroup-Outgroup Perception." In John F. Dovidio and Samuel L. Gaertner (eds.), *Prejudice*, *Discrimination*, *and Racism* (pp. 165–207). New York: Academic Press.

McDonald, Roderick P., and Herbert W. Marsh. (1990). "Choosing a Multivariate Model: Noncentrality and Goodness of Fit." *Psychological Bulletin* 107, 247–255.

Munson, J. Michael. (1984). "Personal Values: Considerations on Their Measurement and Application to Five Areas of Research Inquiry." In Robert E. Pitts, Jr. and Arch G. Woodside (eds.), *Personal Values and Consumer Psychology* (pp. 13–33). Lexington, MA: Lexington Books.

Munson, J. Michael, and Shelby H. McIntyre. (1979). "Developing Practical Procedures for the Measurement of Personal Values in Cross-Cultural Marketing." *Journal of Marketing Research* 16, 48–52.

Novak, Thomas P., and Bruce MacEvoy. (1990). "On Comparing Alternative Segmentation Schemes: The List of Values (LOV) and Values and Life Styles (VALS)." *Journal of Consumer Research* 17, 105–109.

Rokeach, Milton. (1973). *The Nature of Human Values*. New York: Free Press.

Schwartz, Shalom H., and Wolfgang Bilsky. (1990). "Toward a Theory of Universal Content and Structure of Values: Extensions and Cross-Cultural Replications." *Journal of Personality and Social Psychology* 58, 878–891.

Shrum, L. J., and John A. McCarty. (In press). "Issues Involving the Relationships Between Personal Values and Consumer Behavior: Theory, Methodology, and Application." In Lynn R. Kahle and Larry Chiagouris (eds.), *Values*, *Lifestyles*, *and Psychographics*. Hillsdale, NJ: Erlbaum.

Shrum, L. J., John A. McCarty, and Tamara L. Loeffler. (1990). "Individual Differences in Value Stability: Are We Really Tapping True Values?" In Marvin E. Goldberg, Gerald Gorn, and Richard W. Pollay (eds.), *Advances in Consumer Research* (vol. 17) (pp. 609–615). Provo: Association for Consumer Research.

Wyer, Robert S., Jr., Galen V. Bodenhousen, and Theresa F. Gorman. (1985). "Cognitive Mediators of Reactions to Rape." *Journal of Personality and Social Psychology* 48, 324–338.