Call #: HF5415 .A319a

Location: Branson

Item #:

Journal Title: Journal of the Academy of Marketing Science

Volume: 24
Issue: 4
Month/Year: 9 1996
Pages: 366-

Article Author: Herche, Joel,

Article Title: Reversed-Polarity Items and Scale Unidimensionality.

Imprint: EBSCO
Reversed-Polarity Items and Scale Unidimensionality

Joel Herche  
*University of the Pacific*

Brian Engellant  
*University of Evansville*

A standard procedure in the development of multi-item measurement scales is to incorporate reversed-polarity items to control for and/or identify acquiescence response bias. In spite of the broad acceptance of this approach, very little work has been done to evaluate the impact of reversed-polarity items on the dimensionality of scales. This study empirically evaluates the impact of reversed-polarity items on the unidimensionality of several well-known marketing measures. The authors suggest that use of reversed-polarity items may present a substantive problem for marketing scholars because of the resulting degradation of scale unidimensionality. The existence of this phenomenon is confirmed in a multisurvey, multiscale, binational research design. Implications for marketing scale developers and measurement theoreticians are discussed, theoretical bases that might explain the phenomenon are explored, and further research suggestions are presented.

With the growing popularity of multi-item rating scales among marketing academics and practitioners has come the acceptance of “paradigms” or sets of rules for measurement development. These guidelines have long suggested that items with reversed statement polarity (and with reversed coding) should be incorporated into the scales to correct for “yea-saying” or response acquiescence (Heaven 1983; Ray 1979, 1983; Spector 1992).

More recently, the assessment of unidimensionality has been presented as a top priority in scale development (Anderson, Gerbing, and Hunter 1987; Gerbing and Anderson 1988; Hattie 1985), yet little research has addressed whether or how the practice of reversing statement polarity affects scale unidimensionality. Accordingly, this study (1) reviews the research on reversed-polarity items and scale unidimensionality, (2) empirically evaluates the impact of employing reversed-polarity items on the unidimensionality of several measures familiar to marketing researchers, (3) examines literature relating to possible causes for dimensionality degradation, and (4) proposes a research program designed to empirically explore the issues raised.

**RESEARCH FOUNDATIONS**

Scale developers have long been aware of the potential distortion in responses arising from acquiescence (Edwards 1957; Likert 1932), affirmation, or agreement bias, although there is some disagreement as to the severity of the problem (Rorer 1965; Spector 1987). The usual solution offered by contemporary psychometricians is to include both positively and negatively worded items in their measures. For instance, Falchzik and Jolson (1974) found evidence that consumer attitudes and behavior vary in intensity depending on whether the researcher’s state-
ments are phrased positively or negatively. Churchill (1979) prescribed a measurement development approach in which some of the statements would be recast to be positively stated and others recast to be negatively stated to reduce yea-saying or nay-saying tendencies. Spector (1992) also pointed out that varying the direction of questioning minimizes bias produced by response tendencies.

Nowhere is the practice of reversing the polarity of scale items more heartily endorsed than in a stream of research on acquiescence from Ray (1979, 1983) and Heaven (1983). Ray (1983) stated succinctly that it is simply impossible to check unbalanced scales and that their many relationships could be but the consequence of response acquiescence.

Indications of Adverse Consequences

In spite of the aforementioned benefits derived from their application, reversed-polarity items have been presenting problems for marketing scale developers. For example, in an empirical test of the impact of acquiescence, Terborg and Peters (1974) found that differences in scores between reversed- and standard-polarity items appear to occur independent of the effects of acquiescence. Also, Goldsmith and Desborde’s (1991) assessment of an opinion leadership measure revealed that eliminating a negatively worded item from the measure in question would drive internal consistency. Citing Nunnally (1978) and Falthzik and Jolson (1974), they concluded that the renegade item was the only one of the seven items worded in the negative direction. This suggested that other items in the measure needed to be reverse worded to place explicit controls on the direction-of-item wording problem.

Parasuraman, Barry, and Zeithaml (1991), in their response to Carman’s (1990) criticism of their 22-item SERVQUAL scale, “corrected” the polarity of all six reversed-polarity items through rewording. Their justification for the scale revision was based on three criteria targeting the negative-polarity items: (1) larger standard deviations, (2) a perception of potential confusion from awkward wording in a test of executive managers, and (3) poor reliability (Cronbach’s alpha). Interestingly, they did note that dimensionality was an issue in the measure but cited internal consistency as the predominant criterion.

DeVellis (1991) prescribed a cautious approach to the matter. Although recognizing that reversing item polarity addresses the problem of acquiescence, he also saw potential for confusion, especially when completing a long questionnaire. Although no empirical evidence was cited supporting his position, his solution was to be aware of both the acquiescence and confusion problems and to be as clear as possible in the phrasing of questions and instructions. Respondent confusion suggests a nonsystematic variation in responses, reducing the internal consistency of the items but not necessarily affecting their scale dimensionality. Nunnally and Bernstein (1994) went further by suggesting the development of balanced scales that consist of an equal number of positive and negative polarity items. In addition, these authors observed that when items are keyed in different directions, factors tend to arise following the polarity of the items’ keying because endorsing a trait is not the same as rejecting its opposite. When “factors tend to arise,” whether from trait or method sources, the dimensionality of the scale is adversely affected.

The Nonsymmetry of Positive and Negative Items

Other research has reported asymmetrical relationships between positive and negative phrasing of response alternatives. Anderson (1965) and Feldman (1966) both found that negative adjectives are more powerful than positive adjectives in affecting one’s overall evaluation. Jordan (1965) concluded that the custom of finding an arithmetic average of attitude and opinion ratings that includes both positive and negative ratings appeared to be unjustifiable. Also, Kanouse and Hanson (1971) noted that an implicit postulate in measurement development is that there is a single good/bad dimension. They went on to provide evidence that this assumption may be unwarranted.

Differences in the weighting of positive and negative information have been observed recently across a number of research streams including consumer affect (Conway, Bonneville, and DiFazio 1990; Oliver 1993; Westbrook 1987), individual performance (Jaworski and Kohli 1991; Whitehead 1987), purchase involvement (Maheswaran and Meyers-Levy 1990), response to price change (Kalwani, Yin, Rinne, and Sugita 1990), valuation (Hauser, Urban, and Weinberg 1993; Mowen and Mowen 1991), and advertising effectiveness (Pechmann 1992). Although the observed differences vary from item to item and from respondent to respondent, the findings provide ample support for Kanouse and Hanson’s (1971) contention that positively and negatively phrased items do not necessarily lie on opposite ends of the same dimension. If reversed-polarity items do tend to reflect different dimensions, then the measures incorporating them cannot be consistently unidimensional.

The notion of unidimensionality in measurement theory refers to a condition in which a set of indicators share only a single underlying factor (McDonald 1981). More than simply desirable, unidimensionality has been identified as a crucial foundational assumption of measurement theory (Hattie 1985). The effort to obtain unidimensional measures has been regarded as a critical component in theory testing and development research (Anderson et al. 1987). Gerbing and Anderson (1988) articulated an assessment approach and suggested that dimensionality evaluation be incorporated as a critical step in Churchill’s (1979) measurement development paradigm to provide a necessary, but not sufficient, condition for claiming construct validity. Gerbing and Anderson (1988) underscored the fact that reliability measures, including Cronbach’s alpha, mean nothing if the scale is not unidimensional.

A review of the literature suggests that the relationship between statement polarization and measure unidimensionality has not been discussed adequately, much less evaluated empirically. Apart from DeVellis’s (1991) brief...
reference to confusion and Winkler, Kanouse, and Ware's (1983) proposed transformations to improve factor structure interpretability, no cautionary appeals are to be found in the contemporary scale development marketing literature. This is perhaps due to the fact that unidimensionality has only recently been recognized as a necessary prerequisite to construct validity. Accordingly, although the use of reversed-polarity items to control for acquiescence bias has become a widespread practice, there is sufficient indication, as described earlier, to suggest that the practice may have some adverse effects on unidimensionality.

Given the indications presented in the foregoing review, the use of reversed-polarity items in multi-item scales is expected to result in two phenomena with respect to unidimensionality. First, measures with a substantial number of negatively worded items will tend to exhibit a two-dimensional factor structure with the item loadings reflecting the polarity. Second, the introduction of negatively worded items to unipolar measures will cause degeneration in fit when measured against a unidimensional model. Items reversed to "keep respondents honest" will fragment their measures' dimensional characteristics.

**Methodology**

**Approach 1**

To check for the existence of dimensionality effects in measures using reversed-polarity items, three established marketing scales were selected for study. The first was a 24-item measure designed to evaluate salespeople's customer orientation (SOCO). The scale was initially introduced and validated by Saxe and Weitz (1982) and has been used subsequently in several other studies (Michaels and Day 1985; Brown, Coulter, and Widing 1991; Swenson and Herche 1994). Although the number of points used in some administrations apparently underwent minor modifications, evidence of the reliability and construct validity of the SOCO appeared to be strong. Half of the 24 items in the SOCO possess reverse-polarization characteristics.

The second measure included in this study was Spiro and Weitz's (1990) ADAPTS, a 16-item scale developed to assess the degree to which salespeople practice adaptive selling. The construct has been cited in the sales management literature (Weitz, Sujan, and Sujan 1986), and the ADAPTS scale continues to gain the interest of sales force researchers (Badovich and Thompson 1994; Herche, Swenson, and Verbeke forthcoming). Of the 16 items in the ADAPTS, 6 are negatively worded.

The third scale used for assessment in this section was Shimp and Sharma's (1987) 17-item CETSCALE. The measure was designed to assess the perceived morality of purchasing foreign products. The CETSCALE was found to possess nearly equivalent goodness-of-fit indexes across the four regions studied (Shimp and Sharma 1987) and has also shown strong psychometric properties that are robust across cultural settings (Netemeyer, Durvasula, and Lichtenstein 1991). The original CETSCALE does not contain reversed-polarity items.

As shown in Table 1, a multisurvey/binational design was employed to evaluate the selected measures. The SOCO and ADAPTS were administered in two independent surveys of salespeople in the United States and one survey of salespeople in the Netherlands. The questionnaire used in the Dutch survey was subjected to a rigorous back-translation procedure as suggested by Douglas and Craig (1983). The fourth survey involved a nationwide administration of a revised version of the CETSCALE that included seven reversed-polarity items. The revised version of the CETSCALE was pretested on 52 arbitrarily selected individuals for readability, clarity, and reliability before the national sample was drawn. The differences in wording are presented in Table 2.

Principal components factor analysis was then employed as a preliminary step to evaluate the degree to which negatively worded items tended to load onto a factor different from that loaded onto by positively worded items. The negative scale items were expected to load onto separate and distinct dimensions. Confirmatory assessment of the improvement in fit that could be achieved from a two-factor measurement model was subsequently undertaken to reveal the reversed-polarity effects.

**Approach 2**

Following the preceding analyses, a separate field experiment was conducted as part of a scale development effort relating to new product innovation in organizations. In the experiment, 12 indicators (6 positively phrased and 6 negatively phrased) were identified by a group of eight measurement development experts as representing product innovativeness, the construct of interest. All 12 were included in a questionnaire mailed to a national sample of product development executives. A 7-point Likert-type
TABLE 2
Altered Wording of the CETSCALE

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Original Wording</th>
<th>Reversed-Polarity Wording (used in fourth survey)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>American people should always buy American-made products.</td>
<td>American people should not always buy American-made products.</td>
</tr>
<tr>
<td>5</td>
<td>Purchasing foreign-made products is un-American.</td>
<td>Purchasing foreign-made products is not un-American.</td>
</tr>
<tr>
<td>7</td>
<td>A real American should always buy American-made products.</td>
<td>A real American should not always buy American-made products.</td>
</tr>
<tr>
<td>9</td>
<td>It is always best to purchase American products.</td>
<td>It is not always best to purchase American products.</td>
</tr>
<tr>
<td>12</td>
<td>CURBS should be put on all imports.</td>
<td>CURBS should be taken off of all imports.</td>
</tr>
<tr>
<td>14</td>
<td>Foreigners should not be allowed to put their products on our markets.</td>
<td>Foreigners should be allowed to put their products on our markets.</td>
</tr>
<tr>
<td>17</td>
<td>American consumers who purchase foreign products are responsible for putting their fellow Americans out of work.</td>
<td>American consumers who purchase foreign products are not responsible for putting their fellow Americans out of work.</td>
</tr>
</tbody>
</table>

TABLE 3
Scales A and B

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Scale A Item Wording</th>
<th>Scale B Item Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Engineers make a significant contribution to our success.</td>
<td>Engineers have little to do with our product success. (RP)</td>
</tr>
<tr>
<td>2</td>
<td>Consumers inspire our product development program.</td>
<td>Consumers rarely inspire our product development program. (RP)</td>
</tr>
<tr>
<td>3</td>
<td>We set high goals for the number of products we develop.</td>
<td>We set easy goals for the number of new products we develop. (RP)</td>
</tr>
<tr>
<td>4</td>
<td>Our management encourages us to &quot;brainstorm.&quot;</td>
<td>Our management encourages us to &quot;brainstorm.&quot;</td>
</tr>
<tr>
<td>5</td>
<td>We take great pride in our ability to adapt to new technology.</td>
<td>We take great pride in our ability to adapt to new technology.</td>
</tr>
<tr>
<td>6</td>
<td>We strive for greater productivity from our product development efforts each year.</td>
<td>We strive for greater productivity from our product development efforts each year.</td>
</tr>
</tbody>
</table>

NOTE: RP = reversed-polarity item.

response format was employed. Data collection resulted in 142 usable responses (33% response rate).

The data were analyzed to develop a purified scale according to established scale development procedures. Scale A was constructed to comprise the six best-performing items, all of which happened to be positively phrased. The scale was then subjected to confirmatory factor analysis to assess unidimensionality.

Scale A was later revised by recasting three of its items to negative polarity. This revision is referred to as Scale B (see Table 3). Scale B was included in a new survey instrument and was sent to a second independent sample of product development executives, resulting in 200 usable responses (29% response rate). Again, unidimensionality was assessed using confirmatory factor analysis.

Even though scale development procedures were followed precisely and accurately, the unipolar Scale A was expected to present better evidence of unidimensionality than was Scale B. Differential perceptions of negatively worded items were expected to alter the items' meaning and performance as well as to reduce their fit within a unidimensional model of the construct of interest.

RESULTS

The results of an exploratory assessment of factor loading patterns using principal components analysis with varimax rotation of extracted factors revealed a clear and consistent tendency for items to load according to statement polarity. The solutions were constrained to two-factor outcomes to evaluate the degree to which the positive- and negative-polarity items would load on separate orthogonal factors. In Studies 1 and 2, all of the SOCOP's negative polarity items with significant loadings were associated with the same factor, and all of the positive-polarity items loaded on the second factor (a cutoff of .4 was employed per Hair, Anderson, and Tatham 1987). The factor importance was reversed for the Dutch sample, but the pattern remained the same. With the exception of item 12, the loading pattern also held for the ADAPTS. The revised CETSCALE, with 7 of its 17 items negatively worded, also clearly revealed a positive/negative factor structure, although 2 of the 7 reversed items did not load on either factor. This might be explained by the fact that the CETSCALE, although pretested, was revised in a somewhat arbitrary manner, whereas the other measures had been rigorously purified and used for years with the mixed-polarity structure.

The dimensionality of the measures also appeared to be affected by the pattern of factor loadings. For the SOCOP, the first factor accounted for only 20 to 30 percent of variation in the three studies. Although the percentage of variation explained by the first factor was better for the ADAPTS (33% to 36%), the figures still were far below Hair et al.'s (1987) suggested benchmark of 60 percent for a unidimensional measure. The revised CETSCALE appeared to present a stronger claim to unidimensionality; however, a second factor with an eigenvalue of greater than 1 was extracted for it as well.
As explained earlier, unidimensionality assessment was conducted using confirmatory factor analysis for each of the scales with reversed-polarity items (the SOCO, the ADAPTS, and the revised CETSCALE). Two contrasting models consisting of a simple single-factor constraint and a two-factor solution in which the reversed-polarity items were constrained to load onto the second factor were compared for fit characteristics. The outcome revealed that for all of the scales in all of the surveys, the inclusion of a second factor with loading patterns constrained to follow item polarity produced a significant improvement in goodness of fit. Of the 21 fit statistics, 17 were improved by the addition of the second dimension, and the change in the fit was statistically significant ($\alpha = .01$) in all seven tests. Findings related to the CETSCALE indicate the best claim to unidimensionality of the three measures tested because the high correlation (.85) between the two factors suggested that the factors could be collapsed into one. Average interfactor correlations across the three studies were .33 for the SOCO and .59 for the ADAPTS.

Unidimensionality assessment results computed by LISREL 8 for the two innovativeness scales are presented in Table 4. Results provided favorable evidence that the data for Scale A fit a unidimensional model when subjected to confirmatory testing. The $\chi^2/df$ ratio was less than 2.0, the goodness-of-fit and adjusted goodness-of-fit indexes were .94 or greater, and the root mean square residual was less than .05, all established benchmarks for “acceptable” fit (Byrne 1989).

Changing the polarity of scale items was expected to degrade the evidence of unidimensionality in this experiment, and that is certainly consistent with the observed results. Scale B, with three reversed-polarity items, generated weak goodness-of-fit values when subjected to confirmatory testing in a unidimensional structural equation model. The $\chi^2/df$ ratio exceeded 2.0, suggesting that the unidimensional model does not well represent the data presented; the adjusted goodness-of-fit index fell to .90, a substantial difference from the .96 value achieved by Scale A; and the root mean square residual exceeded the .05 guideline, indicating that the differences between observed and estimated input matrices are consistent. Although $p$ values can be misleading in confirmatory factor analysis, the consistency and direction of the change in fit across the two administrations provide substantive evidence of an item-reversal effect. The degradation in fit was deemed sufficient to support the research expectation that the introduction of negatively worded items to a standard-polarity scale would reduce a scale’s performance on unidimensionality tests.

### THEORETICAL DISCUSSION

This research has been designed to empirically evaluate a specific question: does rewording questions so that some items are stated in a negative direction have an impact on measure unidimensionality? The answer to that question, based on data collected for this study, is yes. The next logical question is why. The following review suggests six different theoretical explanations and six variables that may contribute to the incidences of observed asymmetry among response profiles for reversed-polarity item sets. These various perspectives are summarized in Table 5.

### Respondent-Related Variables

#### Expectancy contrast

This view explains asymmetry on the basis of contrast effects in judgment (Helson 1964; Sherif and Sherif 1967). Positivity effects occur when an
observer’s internal reference system causes a stimulus to be perceived as more extreme than objective analysis would indicate. Because most people are exposed to generally positive descriptions of the people, places, and objects in their lives, over time their perceptual anchors may shift toward the positive end of their internalized judgment scales. When they subsequently encounter strongly worded negative stimuli, they will view them more negatively than they would have prior to the positive conditioning. This perspective presumes that the asymmetry will be greater among respondents who have been exposed to larger and longer doses of positive conditioning, causing their internal standards of reference to become skewed. Accordingly, different scale administrations may exhibit varying degrees of asymmetry depending on respondents’ degree of bias.

Memory balance. This perspective suggests that respondents frequently lack ready answers to the questions asked in survey instruments. In lieu of such answers, they run internal tests of hypotheses to determine their responses by calling to mind past behaviors, feelings, and attitudes that support the hypotheses under evaluation (Klayman and Ha 1989). This perspective suggests that when people are asked whether they possess a characteristic, they tend to “favor the null hypothesis” and look for evidence of support rather than refutation. The result is a biased self-conception to the extent that memory evidence of support and refutation is unbalanced. Kunda, Fong, Sanitioso, and Reber (1993) summarized their research supporting this perspective by observing that an analysis of the memories and thoughts brought up by subjects in response to these directional questions alluded that the shifts in self-conceptions showed the subjects’ reliance on positive strategies to test hypotheses about themselves. Subjects searched their memories for instances suggesting that they had the hypothesized characteristic more than they did for instances demonstrating that they did not have it. The researchers found that the biased set of memories elicited in this manner swayed the subjects’ working self-conceptions.

The memory balance perspective predicts that response profiles will be more symmetrical for respondents with memories balanced by both support and refutation. Presumably, respondents who have given the particular question a great deal of cognitive attention because of either past experience or interest will have more balanced memories and more symmetrical response profiles.

Topic sensitivity. This explanation is derived from the attention literature recently summarized by Baron, Logan, Lilly, Inman, and Brennan (1994). The literature suggests that when a subject is negatively aroused, a complex process of appraisal, attribution, and coping is triggered that adversely affects the subject’s performance on ancillary cognitive tasks (such as responding to questionnaires). But Baron et al. found that when the negative stimulus is directly related to the task itself, subjects processed message content more carefully and elaborated on it more fully. Although the study did not deal with survey instrument design, its results pose an interesting perspective on the cause of asymmetry. Negatively phrased items perhaps arouse some sensitive respondents to higher levels of item elaboration, resulting in more complete processing and differential rating. This explanation further suggests that the degree of message elaboration differs among respondents according to their sensitivity to the subject matter.

Item-Related Variables

Frequency-weight. Based on hypothesized differences in the information content of alternate stimuli, this perspective encompasses the explanations offered by corresponding interference theory (Jones and Davis 1965) and the novelty approach (Fiske 1980). As one forms one’s response to the scale item under consideration, cues that are perceived as more informative (negatively phrased items often are considered to be more informative) are weighted more heavily. This explanation suggests that if the positively worded item were reworded to carry more information, then it could be brought into “balance” with the corresponding reversed item. Such action could improve response symmetry.

Range theory. This perspective (Birnbaum 1972; Wyer 1973, 1974) assumes that each scale indicator triggers a restricted range of possible values depending on its degree of distinctiveness. The response is simply the modal value of many possible responses that the subject entertains. Because negative indicators are presumed to be more distinctive than positive indicators, the range of possible responses becomes more narrow and the statistical effect is asymmetry. This perspective suggests that effects may be controlled by balancing the degree of distinctiveness present in response items.

Category diagnosticity. When applied to the problem presented here, category diagnosticity theory (Skowronski and Carlston 1989) assumes that word choices used in scale items offer differing potential to help respondents discriminate between alternative response categories. Category diagnosticity refers to the item’s utility in discriminating among the available response levels or categories, and items depicting extreme behaviors are considered more diagnostic than items containing moderate or neutral behaviors. Because respondents try to maximize their efficiency in making the evaluations necessary to complete a questionnaire exercise, items that are more diagnostic (extreme) are assumed to carry greater influence on respondent impression formation. The results is that subjects confronted with two equal but opposite indicators generally will assign the more diagnostic cue the greater weight in their evaluations. But unlike the previous perspectives, greater weight is not necessarily assigned to the negatively worded item. The greater weight is applied to the more diagnostic indicator. It is conceivable, but unlikely, that the category diagnosticity of the reversed-polarity pair could essentially be the same. If so, then respondents would be expected to apply equal influence to
both items, and response profiles would be symmetrical. (Refer to Skowronsiki and Carlston [1989] for a critique of expectancy contrast, frequency-weight, and range theory perspectives.)

The first item of Shimp and Sharma's (1987) 17-item CETSCALE can serve as an example to illustrate how the six theoretical perspectives apply. Suppose the positively phrased "American people should always buy American-made products" is recast to the negatively phrased "American people should never buy American products" as part of a strategy to control for acquiescence response bias. After reverse coding the negatively phrased item to correct for polarity, statistical results reveal low correlations both to the positively phrased item and to the other items in the measure. The *positive bias perspective* suggests that the word "never" in the reverse item offers greater contrast against the respondents' biased mean point than does the word "always." The *memory balance perspective* posits that respondents have biased sets of memories when testing these items, and *topic sensitivity* theorizes that the negative phrasing arouses a higher level of cognitive processing. The *frequency-weight perspective* suggests that "never buy" is more distinctive, carries greater information content, and consequently is more carefully attended. The *range theory perspective* implies that the negatively phrased item shows less variation because negative items generally are less ambiguous. The *category diagnosticity perspective* hypothesizes that the reversed item is more diagnostic because "never" is the more extreme position, thus earning greater respondent attention.

As indicated in this review, there is broad theoretical support for the idea that reversed-polarity items may behave asymmetrically to standard-polarity items, although the theoretical explanation of this phenomenon from the current literature base remains elusive. When asymmetry is extreme, the reversed-polarity items will tend to behave as distinctive or independent items, unlike their positively worded counterparts. Under these assumptions, asymmetry will adversely affect the measured unidimensionality of the scale and will cause measures that may have been deemed adequate (i.e., were single-polarity construction employed) to test as two dimensional. In view of this, there may be a clear and significant threat to scale unidimensionality when reversed items are included with standard items in the construction of measurement scales. As shown earlier, where there is a threat to scale unidimensionality, there is a threat to scale reliability and validity as well.

**CONCLUSIONS**

Because research in marketing often deals with constructs that are not directly observable and many times are complex in their interpretation, the movement toward multi-item measurement methodologies is important and should be encouraged. As Jacoby (1978) admonished, it is important to know what we do not know, especially when it comes to the quality of our research and the substance and truth of our findings. To interpret research, it is imperative that we demonstrate the validity of the measures we employ.

It has been recognized recently that part of the validation process should involve providing evidence for claiming unidimensionality of all the measures (Anderson et al. 1987; Gerbing and Anderson 1988; Kumar and Dillon 1987). By the same token, response acquiescence or yeasing continues to be a concern for consumer researchers. The conventional solution has been to incorporate reversed-polarity items to force respondents with strong positive or negative attitudes to use both ends of a scale, yet little concern regarding the impact of this practice on the dimensionality of measures has been voiced to date. This is surprising in view of the established and supported findings referenced earlier that "positive" and "negative" are not necessarily perceived to fall on opposite ends of the same dimension. Six theoretical perspectives suggesting six possible explanations have been presented, any one of which may be causing individually, or in combination with others, the response asymmetry observed here that degrades measure unidimensionality.

The primary contribution of this research is to call into question the widely accepted practice of using reversed-polarity items in measurement scales and to highlight the significant potential risk of reducing the unidimensionality of measures. The item polarization decision may be framed as a trade-off between two suboptimal extremes: unidimensional measurement tainted with acquiescence bias and nonbiased measurement tainted by suspect unidimensionality. Because neither scenario is desirable, alternative approaches to measurement development need to be explored.

**Research Extensions**

A series of experimental projects would begin to address some of the research questions raised in this article. The item-related variables discussed earlier can easily be manipulated to reflect varying levels of frequency-weight, distinctiveness, and category diagnosticity. Therefore, this area seems like a logical place to begin. The first step would be to develop items representing high or low levels of one of the variables. A collection of items possessing the appropriate characteristics could be validated by a panel of judges. A balanced (positive/negative) mix of items associated with a given construct would be needed. The degree of degradation in unidimensional model fit or improvement in bidimensional fit could serve as the evaluative criterion in subsequent statistical tests.

The easiest way to design research on respondent-related variables is to separate potential subjects into treatment and control groups based on their answers to a pretest questionnaire designed to classify subjects into response types. Of course, such quasi-experimental approaches can only provide limited evidence of causality.

Experimental approaches to respondent-related assessments could be tested using a balanced $2 \times 2 \times 2$ orthogonal contrast design that would accommodate manipulations of.
each of the three respondent-related variables. The negative arousal of a subject could be pretested through exposure to carefully prepared information on the topic of inquiry and measured using an attitude scale developed specifically for this purpose. Expectancy contrast could be manipulated through the exposure of the test group to a positive environment. Finally, subject memory balance related to the topic of interest could be introduced through the prior presentation of suitable background materials. In all cases, the success of the manipulation would have to be checked through appropriate testing. A control group would be needed for each of the treatments. Testing with orthogonal contrast statements would enable the capture of interaction effects.

Prescriptions to reduce the impact of reversed-polarity items will be different, depending on the results of the proposed research. If item-related variables appear to be driving the effect, then questionnaire design strategies or steps taken at the item generation stage will likely be most effective. The importance of respondent-related variables suggests that the solutions might initially be oriented more toward environmental control factors surrounding the administration of the instrument. Until such research efforts bear fruit, caution is advised when employing scales with mixed-item polarity.

ACKNOWLEDGMENTS

The authors thank the three anonymous reviewers and the editor, Michael J. Swenson of Brigham Young University, Gerald Albaum of the University of Oregon, and Mark Hill of the University of New Orleans for their help in data collection and comments on earlier drafts of this article. The authors also thank Martine Herche for her editorial assistance.

REFERENCES


ABOUT THE AUTHORS

Joel Herche is an assistant professor of marketing and international business at the Eberhardt School of Business, University of the Pacific. He received his Ph.D. from the University of Oregon in 1989. His current research interests include cross-cultural marketing strategy, sales force management, and psychometrics. He is a previous contributor to the Journal of the Academy of Marketing Science and has also been published in the Journal of Personal Selling and Sales Management, International Marketing Review, and European Journal of Marketing, among others.

Brian Engelland is an assistant professor of marketing at the University of Evansville in Evansville, Indiana. Before receiving his D.B.A. from Southern Illinois University at Carbondale, he served in marketing positions for international firms and as marketing chairs for two industry associations. His research interests include product innovation and new product strategy.