

Did You Not Understand The Question Or Not? An Investigation Of Negatively Worded Questions In Survey Research

Jill M. Johnson, Polaris Industries, Inc.

Dennis N. Bristow (E-mail: dbristow@stcloudstate.edu), St. Cloud State University

Kenneth C. Schneider, St. Cloud State University

Abstract

This study compared participant responses to negatively versus positively worded questionnaire statements. The literature suggested that one could expect the wording of questionnaire items (i.e., negatively worded vs. positively worded) to influence participant responses to those scale items. The study consisted of one control group and two experimental groups. The Fashion Consciousness scale (Wilkes 1992; Lumpkin and Darden 1982), a uni-dimensional, all-positive seven-item instrument was adapted for use in the study. Three hypotheses regarding the factor structure and internal reliability of the scale were empirically investigated. Results showed statistically significant differences in the psychometrics of the scale when negative or double negative wording was added to scale items. Managerial implications are discussed.

1. Introduction

In his well-known and highly acclaimed textbook on marketing research, Bill Zikmund offers the following parsimonious definitions of two terms which are key to a meaningful discussion of survey research:

Respondent: The person who verbally answers an interviewer=s questions or provides answers to written questions.
Survey: A method of primary data collection in which information is gathered by communicating with a representative sample of people. (Zikmund 1997, p. 192)

The definitions provided by Zikmund succinctly capture the essence of the survey research process: to contact and receive relevant answers to questions asked of a sample of targeted individuals in order to gain further understanding of some phenomenon. A vast array of companies -- from monoliths like McDonald=s to the corner malt shop -- have embraced the survey research process as a means of discovering what their customers seek in, expect of, and their level of satisfaction with the products and services they buy. The pervasiveness of survey research is evidenced by Kinnear and Taylor=s estimation that over fifty-percent of the general populace in the United States have participated as respondents in one or more survey research projects (1996).

The popularity of survey research is due at least in part to the many advantages offered by the technique. In short, survey research is a versatile, relatively simple, quick, inexpensive, and when done correctly, an accurate and efficient way in which to gather information from one or more market segments. However, as pointed out by a host of authors (i.e., McDaniel and Gates 1993; Aaker, Kumar and Day 1998; Lehmann, Gupta, and Steckel 1998; Churchill 1995), survey research is not without its drawbacks. Hundreds of pages have been devoted to the discussion of the potential errors associated with survey research: random sampling error (i.e., the difference between reported results

based on a random sample versus the result that would have been obtained with a complete census) and, as is discussed more thoroughly in the following section of this paper, systematic error.

Systematic error in survey research is the result or outcome of an imperfection in either the research design itself or in the implementation of the design. The rigorous application of the scientific method by researchers engaging in survey research has helped to reduce the frequency of such errors and at the same time has served to improve the quality of the information gathered via questionnaires. Prior to the implementation stage of the research process, researchers must be cognizant of and actively address a variety of issues when designing the survey instrument. The remainder of this manuscript focuses on a specific source of systematic error: the writing of good survey questions.

2. What Makes A Survey Question “Good”?

Over the years, the survey research process has revealed a plethora of design errors and mistakes that have plagued the survey researcher. Not surprisingly then, experts in marketing research have compiled a quite thorough list of things to >do= and >not do= when writing survey questions. Some of the more common >not to do= items include the avoidance of leading questions, assumptions, and ambiguous words. The work of several researchers has led to a relatively comprehensive list developed by Barnes and Dotson (1989), Alreck and Settle (1995), and adapted by Burns and Bush (1998) that presents a convenient two-faceted categorization of how to write >good= survey questions. The Burns and Bush list of survey questions >shoulds= and >should not= is presented in Table 1.

**Table 1
Shoulds And Should Nots Of Writing Survey Questions***

Shoulds of Writing Survey Questions
<ol style="list-style-type: none"> 1. Questions should be focused on a single issue. 2. Questions should be brief. 3. Questions should be interpreted the same way by all respondents. 4. Questions should use words understood by the respondents. 5. Questions should be grammatically simple to the degree possible.
Should Nots of Writing Survey Questions
<ol style="list-style-type: none"> 1. Questions should not be based upon assumptions. 2. Questions should not go beyond the respondent=s ability or experience. 3. Questions should not use a specific example to represent a general case. 4. Questions should not ask respondents to recall specifics when generalities are likely to be recalled. 5. Questions should not require respondents to guess a generalization. 6. Questions should not ask for details that can not be provided. 7. Questions should not use words that overstate the condition. 8. Questions should not include ambiguous wording. 9. Questions should not be double-barreled. 10. Questions should not lead respondents to particular answer. 11. Questions should not have loaded or value laden wording or phrases.

* Adapted from Burns, Alvin C. and Ronald F. Bush (1998), Marketing Research, Second Edition, Prentice Hall, Upper Saddle River: NJ.

While the list of question dos and don=ts presented in Table 1 is somewhat comprehensive and does serve to make the survey researcher aware of several common problems to avoid when designing a questionnaire, another much discussed and potentially problematic area of questionnaire design is not specifically included in the list: the use of negatively worded questions.

3. Avoiding Negatively Worded Questions

In her book, *A Questionnaire Research: A Practical Guide*, Mildred Patten (1998) cautions against the use of negatively worded questions in survey statements. She writes, "Negatives are easily overlooked and can cause confusion." To illustrate Patten's concern, consider the following negatively worded hypothetical survey question and the meaning of two different participant responses:

1. It is not important to me that my clothes are of the latest style.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

If the respondent circles the number "1" on the scale, indicating that they strongly disagree with the statement, he/she is saying that it is very important that their clothing is of the latest style. On the other hand, if the respondent circles the number "6" on the scale, he/she is telling the researcher that stylish clothing is not an important part of their life. In essence, disagreement with the negatively worded statement means a positive response on the part of the respondent while agreement with the statement means a negative response to the statement. The reader may find such questions and explanations confusing. Imagine the level of confusion likely experienced by the consumer attempting to complete a questionnaire containing similar statements.

4. Double-Negatives

Many writers have cautioned survey researchers against the use of statements containing double negatives (i.e., Hayes 1998; Newman 1995; Payne 1980). There seems to be general consensus in the marketing and consumer behavior literature that respondents typically struggle when trying to formulate an accurate answer to questions worded in the following manner:

1. I usually do not have one or more outfits that are not of the very latest fashion.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

Perhaps the results of a casual pilot of the question with several of the authors' colleagues will provide the reader with a poignant illustration of the problems faced by respondents asked such a question. Most reviewers initially read the question and, after a few moments of reflection, read the question a second time. The reviewers' typical reaction was to then read the question a third time before finally determining which scale point to circle in order to most accurately reflect his/her candid response.

Most readers of this manuscript can probably readily empathize with our colleagues as they struggled to understand and sort through the confusion posed by the above statement. In this case, if the respondent circled the number "1", indicating disagreement with the statement, he/she would be telling the researcher, "I usually *do* have one or more outfits that are *not* of the very latest fashion." Conversely, the survey respondent who strongly agreed with the statement would be telling the researcher, "I usually *do not* have one or more outfits that are *not* of the very latest fashion." Unfortunately, given the degree of confusion experienced by many as they attempt to understand the question, the researcher would be well advised to place little confidence in the respondents' ability to correctly interpret the statement. The astute researcher would subsequently seriously question the meaningfulness of an individual's response to such a statement.

5. Dual Statements

Although the limitations and likely confusion associated with negative and, especially, double negative questions have been addressed in the literature, the strategic use of negatively worded statements in questionnaires has been endorsed by some researchers. In an attempt to avoid directional influence, researchers commonly use dual statements when conducting survey research. That is, it has been suggested that the directionality of questions (i.e., whether the questions are stated negatively or positively) has a direct impact upon an individual=s response to that statement. For example, questionnaires often ask respondents to indicate the degree to which they agree or disagree with a statement. In some cases, the respondent may attempt to answer the question in a manner that pleases or is desired by the researcher (i.e., acquiescence). As shown below, in order to avoid the problem of directional influence, survey researchers frequently develop two different questionnaires: Questionnaire A, featuring the question worded positively and Questionnaire B with the same question worded negatively.

Questionnaire A. Question #1 . I like to shop for clothing.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

It would be quite possible and perhaps even likely that a participant with an acquiescent response style, that is, a respondent with a desire to please the researcher with his/her answer, might decide to agree with the above statement. The participant=s response then becomes not an accurate reflection of his/her like or dislike of shopping for clothing but rather a representation of his/her interpretation of the answer sought by the researcher.

In an attempt to account for the presence of such acquiescent respondents, researchers have often relied on the advantages of a simple random sample and pose the same question, only this time worded negatively, to another sample of respondents. Below is the same question regarding clothing shopping preference, only now presented in a negatively worded format:

Questionnaire B. Question #1 . I do not like to shop for clothing.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

Returning to the previous discussion regarding the possibility of the respondent who seeks to please the researcher, such a respondent would now likely infer that the researcher was looking for respondents who did not like to shop for clothing. The acquiescent respondent would predictably circle the number >6' on the scale above.

By combining the dual question system with the concepts of a simple random and representative sample, the researcher can look for patterns that are indicative of respondents seeking to please the researcher. For example, suppose that the results of data collected using Questionnaire A (featuring the positively worded question) revealed that both female and male respondents enjoyed shopping for clothing (i.e, female mean = 5.6; male mean = 5.2). If the researcher is confident that the respondents to each questionnaire are representative of the population being studied, the researcher could reasonably expect a similar pattern of results from the participants responding to Questionnaire B with the negatively worded question. Now, however, the researcher would expect to find that both females and males consistently disagreed with statement. That is, by consistently circling the numbers >1' or >2' on the scale, the respondents would indicate, much as the respondents in the other group had done by circling >5' or >6' on the scale, that they enjoyed shopping for clothing. If, on the other hand, it was found that the numerical responses of participants receiving Questionnaire B (negatively worded) closely mirrored rather than reversed the responses to Questionnaire A (positively worded), the researcher could reasonably conclude that one of the groups of respondents might have been attempting to meet the expectations of the researcher.

6. Yea-Saying; Nay-Saying

Another problem faced by survey researchers focuses on the observation that some participants may be disinterested in the study, may become bored with the survey, and/or may assume the instrument questions are virtually identical. In such cases, respondents may have a tendency to either agree with virtually every question asked (yea-saying) or to disagree with every question (nay-saying). In order to address this potential problem, some researchers have espoused the alternating format technique. The technique is similar to the dual statement format discussed previously with one major difference. While two separate questionnaires are used with the dual statement technique, the following strategy features negatively and positively worded questions in a single questionnaire. For example, the researcher might include the following items in a questionnaire:

1. I usually have one or more outfits that are of the latest fashion.

<i>Strongly</i>	<i>Disagree</i>	<i>Slightly</i>	<i>Slightly</i>	<i>Agree</i>	<i>Strongly</i>
<i>disagree</i>		<i>disagree</i>	<i>agree</i>		<i>agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

2. It is not important to me that my clothes are of the latest style.

<i>Strongly</i>	<i>Disagree</i>	<i>Slightly</i>	<i>Slightly</i>	<i>Agree</i>	<i>Strongly</i>
<i>disagree</i>		<i>disagree</i>	<i>agree</i>		<i>agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

Suppose that the respondent circled the number > 6' on the scale when answering the first question, thereby indicating that he/she does usually have one or more outfits that are of the latest fashion. If that respondent provided a forthright and candid response to the question, the researcher could reasonably expect the respondent to circle either a >1' or a >2' on the scale when answering the second question. Such a response would indicate that it is important to the respondent that his/her clothes are of the latest style and would be consistent with his/her response to the first question. In this case, the use of one positively worded and one negatively worded statement to measure the respondent=s attitude toward fashion provides the researcher with some assurance as to the veracity of the respondent=s answers. If, on the other hand, the researcher observes that the respondent has circled a >5' or >6' on the scale for each statement, such a response would be indicative of serious inconsistency in the respondent=s attitude and might suggest the presence of >yea-saying= on the part of the respondent.

In order to combat the possibility of, or to at least effectively identify yea-saying/nay saying, researchers commonly rely on the tactic of mixing positively and negatively worded questions (i.e., Need for Cognition Scale, Cacioppo, Petty, and Kao 1984; Materialism Scale, Richins and Dawson 1990; 1992; Self-Esteem Scale, Rosenberg 1965). Such a strategy is often used when researchers are crafting rather lengthy lists of similarly worded questions into one scale.

In summary, a variety of advantages and/or disadvantages have been assigned to the use of negatively worded statements in survey research design. Depending upon which texts or articles one peruses, to which expert one listens or which theory one subscribes, the researcher could embark upon any of at least three or four different avenues concerning the use of negative statements in the survey process. Which design is most appropriate? Should the researcher use negatively worded statements and the dual statement or the alternating format technique? Or is Patten (1998) correct in her observation that negatively worded questions should be avoided entirely in survey research? The remainder of this manuscript is devoted to the discussion of an empirical study designed to test that very question.

7. The Study

The focal point of the research was the comparison of participant responses to negatively versus positively worded questionnaire statements. Based upon the research reviewed, the authors of this manuscript could reasonably expect the wording of questionnaire items (i.e., negatively worded vs. positively worded) to influence participant responses to those scale items. The study consisted of one control group and two experimental groups. The structure of

four items from the fashion consciousness scale (Wilkes 1992; Lumpkin and Darden 1982), a uni-dimensional, all-positive seven-item instrument, was not altered (that is, the four scale items included in the study remained positively worded), for the control group. For the experimental groups the scale items were altered to include either questions including double negatives or to include single negatives. Accordingly, the following hypotheses were espoused and tested:

- H1:** The factor structure of the Fashion Consciousness scale will remain uni-dimensional when all scale items are worded positively.
- H2:** The presence of negatively worded items will adversely effect the internal consistency and/or the factor structure of the FCS.
- H3:** The presence of double negative items will adversely effect the internal consistency and/or the factor structure of the FCS to a greater degree than will negatively worded items.

8. Subjects

A total of 253 students enrolled in variety of courses at a large Midwestern university in the United States participated in the study. Of those participants, 115 were female, 131 were male and 84 percent were Caucasian. The majority of the students (61%) were between the ages of 17 and 21. All participants were enrolled in business or business related courses. Table 2 presents a demographic profile of the subjects.

**Table 2
Demographic Profile Of Participants By Questionnaire Format**

Demographic Variable	Questionnaire Format		
	Positive	Double Negative	Negative
Ethnicity			
African American	0	1 (1.1%)	0
Asian/Pacific Rim	6 (7.3%)	9 (10.3%)	10 (11.9%)
American Indian	1 (1.2%)	0	0
Caucasian	71 (86.6 %)	71 (83.9%)	68 (81.0%)
Hispanic	1 (1.2%)	2 (2.3%)	2 (2.4%)
Other	3 (3.7%)	2 (2.3%)	1 (1.2%)
Age			
less than 17	0	0	0
17-21	57 (69.5%)	50 (57.5%)	47 (56.0%)
22-26	17 (20.7%)	25 (28.7%)	29 (34.5%)
over 26	5 (6.1%)	10 (11.5%)	6 (7.1%)
Gender			
Female	35 (42.7%)	43 (49.4%)	37 (44.0%)
Male	44 (53.7%)	42 (48.3%)	45 (53.6%)

*Note: Due to non response on some items, totals and category responses may differ

9. Research Methodology

In an effort to develop a diverse sample of participants in the study, the data collections were conducted on various days of the week, at various times of day, and in courses from a variety of colleges over a period of several days. Due to the student population sampled in the study, the researchers sought to utilize an existing scale that was appropriate to that population. Accordingly, several questions from the fashion consciousness scale (Wilkes 1992; Lumpkin and Darden 1982) were adapted and included in a survey instrument that was administered to students during regularly scheduled class times. As presented in Tables 3, 4, and 5, three versions of the scale were developed: one

version consisted of alternating negative/positively worded questions; a second version included double negatives; and a third version (from the original FCS) consisted of only positively worded items. The questionnaire items were written into rating statements on a 6-point Likert scale with response categories ranging from strongly disagree (1) to strongly agree (6). Several demographic questions were included in the survey. The three forms of the questionnaire were randomly distributed to student participants during the data collection process. Written and verbal instruction for completion of the questionnaire were provided.

Prior to data collection, a draft of the control group questionnaire was pilot tested with a representative sample of student participants at the previously mentioned university. The pilot suggested that participants had no trouble understanding, comprehending, or formulating candid responses to the scale items. There was no overlap between pilot test participants and participants in the main study.

10. Results

Descriptive statistics, item-to-total correlations, and factor analysis procedures were used to test hypotheses H1, H2, and H3. Each hypothesis is presented below and discussed individually.

H1: The factor structure of the Fashion Consciousness scale will remain uni-dimensional when all scale items are worded positively.

Results: Hypothesis Supported. The scale items used in the study were derived from the work of Lumpkin and Darden (1982) and of Wilkes (1992) in the development of the fashion consciousness scale (Bruner and Hensel 1996). In order to empirically test hypothesis H1, and in a manner consistent with the original FCS, the wording of the four scale items included in the study contained no negatives or double negatives, as in the original FCS. The hypothesis was tested by administering this version of the questionnaire to a group of eighty-two subjects, 35 of whom were female and 44 of whom were male (3 participants chose not to indicate their gender).

Because the previous works of two researchers were consulted and adapted in the current research, the authors of this manuscript strongly believed that the internal reliability of the scale items must be assessed before further data analyses were conducted. Accordingly, the first step in the data analysis was to examine the internal consistency of the scale items.

Following the recommendation of Churchill, Ford, and Walker (1974), the internal consistency of the scale items was assessed by calculating the coefficient alpha for the overall scale. The resultant Cronbach alpha (.846) indicated that the scale items performed adequately in capturing a single construct (Churchill 1979). The observed Cronbach alpha in the study was consistent with that reported by Wilkes (.91) in his work on the FCS (1992).

As a subsequent step in the analysis of the reliability of the FCS scale items, the researchers searched for scale items with relatively low correlations with the total score. Any items with an item-to-total correlation of .35 or less would be eliminated from further scale analyses. No scale items exhibited item-to-total correlations at or below the cut-off point.

The third phase of the scale item analyses consisted of common factor analysis, which was used to check the *a priori* specification of the component structure of the scale. Based upon the literature reviewed and the subsequent hypothesis developed, the researchers expected to find a uni-dimensional factor structure for the FCS items administered to the control group. In order to test that expectation, a common factors procedure with a varimax rotation and no *n*-factor specified was used to further examine the structure of the control group FCS. As seen in Table 3, the common factors procedure resulted in the extraction of the expected single-factor structure.

Table 3
Factor Analysis And Internal Reliability:
Fashion Consciousness Scale C Positively Worded Items

1. I usually have one or more outfits that are of the very latest fashion
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6

2. It is important to me that my clothes are of the latest fashion
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6

3. Dressing smartly is an important part of my life
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6

4. I like to shop for clothing
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6

Reliability Analysis C Fashion Consciousness (FC) Scale						
Scale Item*	Scale Item Mean	Scale Item Std. Dev.	Item-to-Total Correlation	Alpha if Item Deleted	Factor Loadings	
					1	2
1.	4.21	1.56	.728	.784	.867	-
2.	3.54	1.45	.783	.762	.899	-
3.	4.16	1.31	.694	.804	.835	-
4.	3.97	1.67	.561	.865	.728	-
Overall Cronbach=s Alpha = .846					Eigenvalue 1 = 2.78 Variance Explained = 69.681%	

The items in each factor were also examined for the significance level of the correlation between the original variable and its factor. Following the recommendations of Churchill and his colleagues (1974), the authors determined that those items with a factor loading of less than .35 would be eliminated from the scale. No such items were discovered. Table 3 presents the final scale structure and factor loadings after factor analysis. Based upon the reported findings, the authors determined that the control group version of the FCS exhibited adequate internal reliability and subsequently concluded that hypothesis H1 was supported.

H2: The presence of negatively worded items will adversely effect the internal consistency and/or the factor structure of the FCS.

Results: Hypothesis Supported. Hypothesis H2 was developed in order to empirically investigate Patten=s (1998) contention that the use of negatively worded questions are likely to confuse survey participants. Eighty-two students (35 female; 44 male; 3 no gender response) completed the FCS containing negatively worded questions. As in the testing of the previous hypotheses, due to the revisions made to the original FCS items, the recommendations of Churchill, Ford, and Walker were followed and the internal reliability of the scale items was tested. The resultant Cronbach alpha (.665) suggests problems within the revised scale. That is, the FCS consisting of positively worded questions demonstrated a Cronbach alpha of .846, suggesting that the scale items performed adequately in measuring a

single construct. When those same scale items were revised to include two negatively phrased items, the observed alpha dropped to .665, indicating that the revised FCS performed less adequately in measuring the construct of fashion consciousness. In short, the internal reliability of the FCS suffered due to the addition of the negatively worded items.

As in the testing of the previous hypotheses in the study, a common factors procedure with a varimax rotation and no specified *n*-factor was implemented to investigate the structure of the revised FCS. While the analysis did reveal a single factor structure for the FCS, as compared to the original FCS scale analyses, substantial differences were observed in the item-to-total correlations observed. As can be seen in Table 4, the item-to-total correlations for the items included in the negatively revised FCS ranged from a low of .429 to a high of .473 with an eigenvalue of 2.02. By contrast, table 3 shows item-to-total correlations ranging from a low of .561 to a high of .783 with an eigenvalue of 2.78 for the original FCS items. Based upon the final scale structure of the negatively revised FCS (see Table 4), the authors concluded that hypothesis H2 was supported.

Table 4.
Factor Analysis And Internal Reliability:
Fashion Consciousness Scale C Negatively Worded Items

- 1. I usually have one or more outfits that are of the very latest fashion
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6
- 2. It is not important to me that my clothes are of the latest style
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6
- 3. Dressing smartly is an important part of my life
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6
- 4. I do not like to shop for clothing
Strongly disagree *Disagree* *Slightly disagree* *Slightly agree* *Agree* *Strongly agree*
 1-----2-----3-----4-----5-----6

Reliability Analysis C Fashion Consciousness (FC) Scale						
Scale Item*	Scale Item Mean	Scale Item Std. Dev.	Item-to-Total Correlation	Alpha if Item Deleted	Factor Loadings	
					1	2
1.	4.32	1.37	.436	.605	.726	-
2.	3.72	1.28	.472	.581	.724	-
3.	4.37	1.08	.473	.591	.703	-
4.	2.86	1.52	.429	.618	.693	-
Overall Cronbach=s Alpha = .665		Eigenvalue 1 = 2.02 Variance Explained = 50.596%				

H3: The presence of double negative items will adversely effect the internal consistency and/or the factor structure of the FCS to a greater degree than will negatively worded items.

Result: Hypothesis Supported. Based upon the literature reviewed, the authors predicted that the inclusion of scale items featuring double negatives would significantly influence the internal reliability and the factor structure of the FCS. Once again, due to revisions made to the original FCS items (i.e., the addition of double negatives), the authors began the testing of hypothesis H3 by examining the internal consistency of the scale. Eighty-seven participants (43 female; 42 male; 2 no gender response) completed the questionnaire containing double negatives.

The observed coefficient alpha for the overall FCS containing double negatives was .260, with item-to-total correlations ranging from .034 to .202. Such correlations and Cronbach alpha provide strong evidence that the revised FCS items are not internally consistent and may have failed to capture a single construct (Churchill 1979). As a subsequent analysis of the scale items, the authors again used common factor analysis to examine the component structure of the scale. The authors expected to find a lower Eigenvalue for the FCS containing double negatives and, possibly, even a deterioration of the one-factor structure. As in the testing of hypothesis H1, a common factors procedure with a varimax rotation and no *n*-factor specified was used to further examine the structure scale. As seen in Table 5, the factor analysis revealed a two factor structure for the double negative FCS.

Table 5
Factor Analysis And Internal Reliability:
Fashion Consciousness Scale C Double Negatives Included

1. I usually do not have one or more outfits that are not of the very latest fashion.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

2. It is important to me that my clothes are of the latest style.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

3. Not dressing smartly is not an important part of my life.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

4. I like to shop for clothing.

<i>Strongly disagree</i>	<i>Disagree</i>	<i>Slightly disagree</i>	<i>Slightly agree</i>	<i>Agree</i>	<i>Strongly agree</i>
1-----	2-----	3-----	4-----	5-----	6-----

Reliability Analysis C Fashion Consciousness (FC) Scale						
Scale Item*	Scale Item Mean	Scale Item Std. Dev.	Item-to-Total Correlation	Alpha if Item Deleted	Factor Loadings	
					1	2
1.	3.24	1.43	.160	.167	-.287	.771
2.	3.47	1.29	.202	.119	.724	.404
3.	3.04	1.32	.034	.320	-.523	.612
4.	4.31	1.46	.130	.208	.776	.320
Overall Cronbach=s Alpha = .260		Eigenvalue 1 = 1.42 Variance Explained = 35.69%			Eigenvalue 2 = 1.28 Variance Explained = 32.215%	

Based upon the results of the factor analysis and the lack of internal reliability of the revised scale items, the authors concluded that hypothesis H3 was supported.

11. Implications And Extensions

The primary implication of these findings is rather obvious and more than a little disturbing:

The results of the study strongly suggest that it is untrue that negatively worded questions do not adversely affect the pattern of responses to a survey question. Furthermore, it would be inappropriate to conclude that the presence of double negatives does not significantly increase the adverse impact. Researchers must not overlook this conclusion; doing so would not fail to limit the usefulness of market surveys.

The reader who successfully navigates the preceding minefield of negative-laden levity will see through to the core implication of the study: negatively worded questions and, especially, questions with double negatives, confuse many survey respondents. The response pattern obtained with a negatively worded question is not necessarily the inverse of that obtained with the identical question worded positively. To an even greater extent, the response pattern to a double negative question does not parallel that obtained with the identical question worded positively, as it should if all respondents correctly adjust for the double negative.

All of this difficulty is exacerbated when the question is designed to gauge an attitude, opinion or perception, wherein the allowable responses typically follow a scale from negative to positive (e.g., disagree to agree, bad to good) or vice versa (e.g., helpful to not helpful, important to unimportant). In such a case the respondent must correctly deduce toward which end of the scale his or her response should be placed, which is partly a function of the attitude or opinion and partly a function of the number of negatives that need to be waded through in understanding and interpreting the question. The research reported here suggests that negatives in the wording do impact the process of interpreting the questions, leading at least some respondents to misinterpret how to respond and, thus, reducing or destroying the usefulness of the question or series of questions.

This tendency for negatively phrased questions to confuse respondents is especially troubling given the time honored practice by marketing researchers to purposely introduce negatives into survey questions to combat directional influence, respondent acquiescence, boredom or other purportedly undesirable outcomes of including only positive versions of questions. While the researchers here readily stipulate the seriousness of these other sorts of problems, this research strongly suggests that avoiding them by introducing negatives to survey questions comes at a potentially considerable price. That is, solving one set of problems by using negative questions may give rise to a different set of problems involving the ability of respondents to correctly sort through and uncover the intended directionality of a question.

There seems, then, to be two key research issues to address. First, do the results reported here occur generally, or were they study specific. In the present study, the reliability of one particular four-item scale, the Fashion Consciousness Scale, was considerably reduced when the directionality of some scale items was reversed, and was completely destroyed (along with the dimensionality of the scale) when the directionality of some scale items was doubly reversed. It remains to be seen through replication whether the same happens with other scales administered to other sample groups. The authors fear it might. Also, the present study was designed to a specific set of measures of question usefulness; to wit, the impact of negativity in question wording on the internal consistency and factor structure or dimensionality of a multi-item scale. Perhaps other researchers will envision alternative ways to operationally define a question's usefulness.

Second, should replication find it common that respondents are confused by and, therefore, misinterpret negatively worded questions -- or those with double negatives, a more important research issue needs to be addressed. Simply put, are the systematic errors introduced by using negatives in crafting survey questions less serious or more serious than the systematic errors controlled by using negatives? If less serious, than this study becomes an interesting but not terribly important one. If the other way around, however, this study sounds an ominous alarm. Either way, the stakes

are insufficiently unimportant that the marketing research profession must not be satisfied doing nothing further to address the concerns raised (sorry, reader). 📖

References

1. Aaker, David A., V. Kumar, and George S. Day (1998), *Marketing Research*, Sixth Edition, John Wiley and Sons, Inc., New York: NY
2. Alreck, Pamela L. and Robert B. Settle (1995), *The Survey Research Handbook*, pp. 88-99. Richard D. Irwin, Homewood: IL
3. Barnes, James H. and Michael J. Dotson (1989), "The Effect of Mixed Grammar Chains on Response to Survey Questions", *Journal of Marketing Research*, November, 468-472.
4. Bruner, Gordon C. II, and Paul J. Hensel (1996), *Marketing Scales Handbook: A Compilation of Multi-Item Measures*, Vol. II, pp. 283-285. American Marketing Association, Chicago: IL
5. Burns, Alvin C. and Ronald F. Bush (1998), *Marketing Research*, Second Edition, Prentice Hall, Upper Saddle River: NJ
6. Cacioppo, John T., Richard E. Petty, and Chuan Feng Kao (1984), "The Efficient Assessment of Need for Cognition", *Journal of Personality Assessment*, 48 (3), 306, 307.
7. Churchill, Gilbert A. Jr., Neil M. Ford, and Orville C. Walker Jr. (1974), "Measuring the Job Satisfaction of Industrial Salesman", *Journal of Marketing Research*, 11, August, 254-260.
8. Churchill, Gilbert A. Jr. (1995), *Marketing Research: Methodological Foundations*, Sixth Edition, The Dryden Press, Harcourt Brace College Publishers, New York: NY
9. Churchill, Gilbert A. Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs", *Journal of Marketing Research*, 16, February, 64-73.
10. Hayes, Bob E. (1998), *Measuring Customer Satisfaction: Survey Design, Use, and Statistical Analysis Methods*, Second Edition, ASQC Quality Press, Milwaukee: WI
11. Kinneer, Thomas C. and James R. Taylor (1996), *Marketing Research: An Applied Approach*, Fifth Edition, McGraw-Hill, Inc., New York: NY
12. Lehmann, Donald R., Sunil Gupta, and Joel H. Steckel (1998), *Marketing Research*, Addison-Wesley, New York: NY
13. Lumpkin, James R. and William R. Darden (1982), "Relating Television Preference Viewing to Shopping Orientations, Lifestyles, and Demographics", *Journal of Advertising*, 11, (4), 56-67.
14. McDaniel, Carl Jr., and Roger Gates (1999), *Contemporary Marketing Research*, Fourth Edition, South-Western College Publishing, New York: NY
15. Newman, Lynn (1995), "That's A Good Question", *American Demographics*, June, pp. 10-15.
16. Patten, Mildred L. (1998), *Questionnaire Research*, Pycszak Publishing, Los Angeles: CA
17. Payne, Stanley L. (1980), *The Art of Asking Questions*, Princeton University Press, Princeton: NJ
18. Richins, Marsha L. and Scott Dawson (1990), "A Preliminary Report of Scale Development", in *Advances in Consumer Research*, Vol. 17, Marvin E. Goldberg, Gerald Gorn, and Richard W. Pollay, eds. Provo, UT: Association for Consumer Research, 169-175.
19. Rosenberg, Morris (1965), *Society and the Adolescent Self-Image*, Princeton University Press, Princeton: NJ
20. Wilkes, Robert E. (1992), "A Structural Modeling Approach to the Measurement and Meaning of Cognitive Age", *Journal of Consumer Research*, 19 (September), 292-301.
21. Zikmund, William G. (1997), *Exploring Marketing Research*, Sixth Edition, The Dryden Press, Harcourt Brace College Publishers, New York: NY