

## THE BIASING EFFECTS OF SCALE-CHECKING STYLES ON RESPONSE TO A LIKERT SCALE

Hershey H. Friedman, Brooklyn College, CUNY  
Paul J. Herskovitz, College of Staten Island, CUNY  
Simcha Pollack, St. John's University

Hershey H. Friedman, Department of Economics, Brooklyn College  
Brooklyn, New York 11210

**KEY WORDS:** Marketing, Survey, Questionnaire, Rating scale, Likert scale, Bias

### ABSTRACT

A study was conducted with a sample of 208 subjects to determine whether the wording of the response categories used in a Likert scale would cause differences among responses to two types of Likert scales: a scale in which all the response categories were worded "strongly agree," "agree," "undecided," "disagree," and "strongly disagree" and a scale in which all the response categories were worded "strongly disagree," "disagree," "undecided," "agree," and "strongly agree." The scale consisted of ten items designed to rate student's attitudes toward their college. Results indicated that the responses to the two scales were significantly different. The scale with the "strongly agree" response category on the left side resulted in greater degree of agreement than the scale with the "strongly disagree" on the left side. This indicates the existence of a bias towards the left side of the scale.

### INTRODUCTION

Attitude scales are widely used by researchers to measure people's attitudes towards a variety of stimuli such as products, services, institutions, occupations, etc. The Likert scale is easy to use, merely requiring subjects to indicate their extent of agreement or disagreement with each of several statements. For example, if a statement is made about the quality of IBM computers, the subject would be asked to check either "strongly agree," "agree," "neither agree nor disagree" (or "undecided"), "disagree," "strongly disagree." These responses are then given values, typically from 1 to 5. The Likert scale is then analyzed in one of two ways: either on an item-by-item basis (profile analysis), or by summing the numerical value of the responses to each item thereby yielding one score per subject for the whole attitude scale (aggregate analysis). If the latter approach is used, the researcher must reverse score unfavorable items. Thus, strongly agreeing with a favorable statement results in the same score as strongly disagreeing with an unfavorable statement.

Similarly, strongly agreeing with an unfavorable statement yields the same score as strongly disagreeing with a favorable statement. For a valuable discussion regarding the construction of Likert scales the reader is referred to Lundstrom and Lamont (1976) and Churchill(1979).

One major decision involved in the construction of a Likert scale asking subjects for their extent of agreement/disagreement, concerns the relative number of favorable items and unfavorable items to include in the scale. Often, one sees Likert scales with all or almost all of the items worded in a favorable way. A study by Friedman (1988) found that a scale consisting of only favorable items (e.g., "\_\_\_\_ College has an excellent reputation") produced results that were significantly different than a scale consisting of unfavorable items (e.g., "\_\_\_\_ College has a terrible reputation.")

Belson (1966), studying the response-order effects of rating scales, found the existence of a bias towards the response category that is listed first. Using a written questionnaire containing five different kinds of rating scales -- satisfaction, agreement, approval, interest, and liking -- Belson compared responses to the traditional order of scale presentation (in which the most positive response categories, such as "very satisfied" or "strongly agree," are listed first) with responses to rating scales consisting of the same response categories but in reverse order (in which the negative response options, such as "very dissatisfied" or "strongly disagree," are presented first). He found that the reversed order of presentation gave rise to a "negative shift" whereby the negative end of the scale received more responses than when presented in the traditional manner. However, Belson suggests further follow-up research to determine whether the reversal phenomenon will occur with horizontal verbal rating scales (i.e., scales in which all the response categories are presented on one line rather than vertically listing each option on a separate line.)

Belson's findings were supported by Payne (1972) and Carp (1974). Payne's (1972) study involved mail questionnaires sent to telephone subscribers asking them to rate various aspects of

telephone service. He tested various kinds of rating scales but did not examine Likert-type strongly agree/strongly disagree scales. Carp (1974) investigated response-order effects by conducting home interviews in which respondents were handed cards containing the response categories which were read aloud by the interviewer. She found that both the negative and positive ends of the scale elicited more responses when presented first and resulted in fewer responses when presented last. This effect was observed for attitudinal questions (i.e., degree of satisfaction) but not for factual questions. However, Powers et al.(1977) failed to find response-order effects when re-analyzing Carp's data. They also conducted their own study to determine the effects of having interviewers reverse the serial order of various types of rating scales. Their comparison of the responses to the two types of scales showed no statistically significant differences.

Holmes (1974), using a sample of 240 beer drinkers, comparing bipolar scales going from left to right with the same scale going from right to left (e.g., warm/cold vs. cold/warm), found a definite bias towards the left side of the scale.

The purpose of the current study was to determine whether there would be any differences between the responses to two types of Likert scales: a scale in which the response categories were worded from left to right as "strongly agree", "agree," "undecided," "disagree," and "strongly disagree" (SA/SD) and a scale in which all the response categories were worded from left to right as "strongly disagree," "undecided," "agree," and "strongly agree" (SD/SA). Both scales consisted of ten items designed to rate student's attitudes toward their college.

## METHOD

A sample 208 undergraduate students at a large urban college was asked to rate the college. Subjects were randomly assigned to one of two questionnaires. The only difference between the two questionnaires was in the order of the response categories to a ten-item Likert scale. In one questionnaire the direction of the five response categories (left to right) was "strongly agree"... "strongly disagree" and in the other the direction of the response categories (left to right) was "strongly disagree"... "strongly agree."

Students indicated the extent of their agreement/disagreement, on a five-point scale, to each of ten items. These items were:

- College faculty is extremely unqualified
- College courses are useful
- College has an excellent reputation
- College has excellent computer facilities
- College has terrible library facilities
- College faculty members are not at all knowledgeable in their areas of specialty
- College offers a broad selection of courses
- College faculty is extremely helpful in working with students
- College staff is not at all helpful
- Overall,  College is a terrible college

A value of 1 was assigned to "strongly agree," a value of 2 to "agree," ..., and a value of 5 to "strongly disagree." This scoring system was used for both questionnaires (SA/SD and SD/SA).

## RESULTS

The experimental manipulation was considered successful since the mean responses to three questions common to both questionnaires -- involving ratings of the college, core courses, and student social life -- were all statistically equivalent (no individual *t*-value was greater than .56, *p* > .50). These three questions utilized a 7-point rating scale which used the following descriptors: excellent, very good, good, average, poor, very poor, and terrible. Values of one through seven were assigned to these descriptors in order to compute means.

Table 1 displays the means, standard deviations, and univariate *t*-values for each of the ten items in the experiment. If the direction of the response categories does not make a difference, then the means for the two groups should all be statistically equivalent. A multivariate analysis of variance (MANOVA) on the ten-item Likert scale resulted in a Wilks' Lambda statistic of 0.91, which is approximately by an F-statistic (d.f. = 10, 197) of 2.04. This was significant at *p* < .05, indicating that the vectors of means for the two groups were significantly different. Thus, an analysis of responses to the Likert scale from a multivariate approach (i.e., considering the responses to all ten items simultaneously) indicated that there was a significant difference in how subjects responded.

Three of the ten univariate F-values were significant at the .05 level. For item 2 (courses are useful), item 3 ("college has excellent reputation"), and item 7 ("college offers a broad selection of

courses"), there was a statistically significant difference between the means. In each of these cases, the mean obtained from the SA/SD scale was lower than the mean from the SD/SA scale. This demonstrated that subjects were more likely to agree with the above statements when the response category of "strongly agree" was on the left side (the beginning of the scale) rather than when it appeared on the right side of the page (the end of the scale).

This bias towards the left side of the scale appeared to manifest itself when the items were worded favorably. Of the five items that were worded favorably (items 2, 3, 4, 7, and 8), three showed significant differences between the SA/SD and SD/SA response categories. Among the items that were worded unfavorably, there were no significant differences. It should be noted that the students, for the most part, held positive attitudes towards their college. It seems that with favorable items, subjects had a greater tendency to check the left side of the scale to show their agreement than to check the right side of the scale to indicate agreement. With the unfavorable items -- items with which most of the students were going to disagree -- it apparently did not seem to make a difference whether the "disagree" categories were placed on the right or left side of the scale.

## DISCUSSION

This study demonstrates that using a Likert scale with the response categories ordered from "strongly agree" to "strongly disagree" will produce different results than a scale with the same items but with the response categories ordered in reverse from "strongly disagree" to "strongly agree." This indicates the existence of a bias towards the response category that is listed first (i.e. on the left side) on the horizontally presented agreement scale, consistent with Holmes (1974).

In addition, this effect manifested itself in the current study when subjects were presented with a favorable statement. This effect did not appear for any of the negative statements. One possible explanation for this may be that, in this particular study, students clearly held positive attitudes towards their college. Thus, their responses to unfavorably worded items required some degree of active participation considering the effort necessary to overcome any yea-saying response tendency. Since the responses to the unfavorably worded items did not require this kind of cognitive effort, the bias towards

the left side of the scale would be more in evidence. More research is necessary in order to confirm or reject this hypothesis.

## REFERENCES

- Belson, W.A. (1966), "The Effects of Reversing the Presentation Order of Verbal Rating Scales," *Journal of Advertising Research*, 6 (December), 30-37.
- Carp, F.M. (1974), "Position Effects on Interview Responses," *Journal of Gerontology*, 29 (September), 581-587.
- Churchill, G., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.
- Friedman, H.H. (1988), "The Effects of Positive and Negative Wording on Responses to a Likert Scale," *Applied Marketing Research*, 28 (Fall), 17-22.
- Holmes, C. (1974), "A Statistical Evaluation of Rating Scales," *Journal of the Market Research Society*, 16 (April), 87-107.
- Lundstrom, W.J. and L.M. Lamont (1976), "The Development of a Scale to Measure Consumer Discontent," *Journal of Marketing Research*, 13 (November), 373-381.
- Payne, J.D. (1972), "The Effects of Reversing the Order of Verbal Rating Scales in a Postal Survey," *Journal of the Market Research Society*, 14 (January), 30-44.
- Powers, E.A., P.Morrow, W.J. Goudy, and P. Keith (1977), "Serial Order Preference in Survey Research," *Public Opinion Quarterly*, 41 (Spring), 80-85.

**TABLE 1**  
**Means, Standard Deviations and Univariate t-tests for the Ten Items**

<u>Item</u>	<u>SA/SD</u> (n=104)		<u>SD/SA</u> (n=104)		<u>t-value</u>	<u>prob.</u>
	<u>Mean</u>	<u>std.dev.</u>	<u>Mean</u>	<u>std.dev.</u>		
1. faculty qualification	3.57	.90	3.64	.94	-.60	.55
2. courses useful	2.10	.60	2.31	.75	-2.24	.03
3. reputation	2.14	.78	2.42	.95	-2.31	.02
4. computer	2.50	.78	2.55	.70	-.47	.64
5. library	3.72	.91	3.58	.88	1.16	.25
6. faculty knowledge	3.69	.88	3.78	.89	-.70	.48
7. course selection	2.20	.82	2.54	1.00	-2.65	.01
8. faculty helpfulness	2.71	.92	2.87	1.12	-1.08	.28
9. staff helpfulness	3.47	.91	3.36	.98	.88	.38
10. overall	4.28	.63	4.13	.78	1.47	.14