SEQUENCE, POLE AND CONTEXT EFFECTS IN SIMILARITIES RATINGS

John R. Dickinson, University of Windsor

Abstract

Numerous potential sources of bias relating to questionnaire layout have been identified. Using a multivariate approach, this research investigates the presence of three types of order bias in the collection of similarity ratings. Ordinal position and pole orientation effects were generally not significant while contextual contamination was consistently present.

Introduction

The possible existence of response biasing effects deriving from various elements of questionnaire design has long been recognized. Such effects may be of a variety of sorts and stem from processes that to date have not been placed in an orderly comprehensive theoretical framework. The particular general situation of interest here is where the respondent is presented with a series of similar rating scales for the purpose of measuring several stimulus objects on a single dimension, one stimulus object on several dimensions, or other variations. Specifically, this study concerns the rating of pairs of stimuli with regard to their "similarity," no particular dimension being specified. Data gathered in this manner commonly form the basis for input to multidimensional scaling algorithms (Green and Rao, 1972; Green and Carmone, 1970).

Within the context of unidimensional scaling tasks researchers have postulated and/or investigated several sources of biasing effects. Ferber (1952) found that rating high income professions first induced respondents to set a more strict rating standard for subsequently listed occupations. In a "select x of n" task, Campbell and Mohr (1950) found no effect of ordinal position, while Becker (1954) found that the proportion of times a stimulus was selected decreased steadily with later ordinal positions. Landon (1971) found significant contextual contamination in the application of a semantic differential scale. Other authors have cited a tendency to select responses at the beginning of a line due to carelessness (Vernon, 1939), an apparent tendency to select items at the extreme positions in a list (Payne, 1951), an anchoring effect whereby early responses provide comparative benchmarks for subsequent responses (Sherif, Sherif and Nebergall, 1965), and a motive for internal consistency (McGuire, 1960), as potentially biasing factors. In the somewhat different context of paired comparison taste tests, Greenburg (1958, 1963) reports a significant order of presentation effect, while Day (1969) provides evidence that this effect does not necessarily occur.

In the more problematic context of collecting similarities judgments for multidimensional scaling, wherein alternative measurement approaches have received considerable attention, only one study of order effects has been published. Jain and Pinson (1976), based on a pilot experiment, found that INDSCAL solutions were insensitive to the order of presentation of 56 pairs of cities for rating.

Ordinal position per se would seem to motivate only a single general effect. Namely, due to fatigue and/or waning interest, subjects may respond less thoughtfully, i.e., more randomly, to later items. Alternatively, subjects may employ heuristics yielding systematic (at the individual level) but invalid responses. Both the fatigue-interest and non-common heuristic influences would be reflected in a larger variance in ratings across subjects. However, there seems to be no reason to think that subjects would, in aggregate, systematically rate later items higher or lower than earlier items.

Pole orientation has received very little empirical study and theoretical conflicts between primacy versus recency effects do not suggest that subjects favor the left or right hand portions of a scale.

The presence of contextual contamination is widely, though not universally (Osgood, 1957), recognized. Results of two empirical studies (Ferber, 1952; Landon, 1971) indicate an indirect relationship whereby higher earlier ratings lead to lower subsequent ratings and vice versa.

The purpose of this study is to investigate the presence of sequence, pole, and context effects in similarities ratings and to describe the nature of these effects if present.

Methodology

Data Collection

The stimulus objects utilized in this study were five adjectives (yielding ten stimulus pairs) selected from a list of fifty originally compiled by Myers and Warner (1968). The adjectives selected had low variances on a favorableness dimension as evidenced by the Myers and Warner study as well as results of a pretest of some 158 subjects for this study. The adjective stimuli were also drawn at varying intervals on the favorableness dimension. The intended effects of these selection criteria were to mitigate variance in responses attributable to true differences in subjects' perceptions of the stimuli and to evoke a variety of responses across stimuli (pairs), at least on the single obvious dimension of favorableness. As this study addresses similarities data which are presumably multidimensional and not necessarily of common structure or space across subjects, the intended effects of the selection criteria may be only partially achieved. The ten adjective pairs are presented in Table 1.

Subjects for this research were second and third year undergraduate business students. While appropriate in the sense of constituting an a priori homogeneous experimental group, use of student subjects may limit the generalization of results to other populations. It is possible that students are more analytical and more accustomed to abstract scenarios than are members of other market segments. Consequently, they may be less influenced by the questionnaire design factors investigated herein. Nevertheless, the relative effects of these factors are not necessarily altered and any

TABLE 1
ADJECTIVE PAIRS USED AS STIMULUS OBJECTS

Adjective Pair	Adjectives
1	terrific, exceptionally good
2	terrific, extremely good
3	terrific, remarkably good
4	terrific, quite good
5	exceptionally good, extremely good
6	exceptionally good, remarkably good
7	exceptionally good, quite good
8	extremely good, remarkably good
9	extremely good, quite good
10	remarkably good, quite good

effects found to be significant here are by similar reasoning presumably more pronounced in non-student groups. A field study itilizing business managers is currently underway to corroborate the results reported here.

Printed instructions directed the subjects to "... rate each of these pairs according to how similar the two adjectives are. The key consideration is how much alike are the two adjectives in each pair in terms of overall similarity. If you feel the two adjectives are very similar, you should give the pair a fairly high rating toward the 'more similar' end of the scale. And, of course, if the two adjectives are not so similar you should rate the pair toward the 'less similar' end." The single page of instructions was followed by a second page containing all ten possible pairings of the five adjectives. Thus, the entire set of adjective pairs was available at once to the subjects.

The rating instrument was a nine-point bipolar "less similar - more similar" semantic differential type scale with the corresponding scale value printed under each line segment. For each subject the basic order of presentation (i.e., sequence) of the ten adjective pairs was randomized as were the rating scale poles, and the order of the two adjectives comprising each pair.

Approximately two weeks after the initial phase data collection, a retest of the same subjects was carried out under identical conditions. Seventy-three usable data sets were obtained in the first data collection phase and 55 of these subjects completed the retest.

Operational Hypotheses

The theoretical considerations discussed earlier give rise to the following operational hypotheses.

H1: Sequence Effect - The ordinal position of a given adjective pair does not affect the level of its similarity rating. No a priori rationale suggests that subjects systematically rate stimuli appearing later either higher or lower than stimuli appearing earlier.

H2: Pole Effect - Ratings of a given adjective pair will be higher when the "more similar" (higher score) pole appears at the left and lower when the "less similar" pole appears at the left. A modicum of theory and empirical evidence suggests that subjects will tend to select ratings in the earlier or left-hand portion of the scale.

H3: Context Effect - The higher the ratings of adjective pairs appearing before a given adjective pair the lower the rating of the given adjective pair. It is theorized that the level of the anchoring point framework established by earlier ratings will inversely affect subsequent ratings.

This hypothesized context effect is predicated on the idea of earlier ratings establishing a framework of anchoring points. Presumably the more ratings or anchoring points making up this framework, the greater its influence. Thus, ordinal position is seen to interact with the context effect, the latter increasing as a direct function of the former.

Analysis

A multiple regression analysis was utilized to test the three hypotheses addressing the three different types of effects. Each adjective pair was analyzed separately for both the test and retest phases yielding 20 distinct analyses. The sequence effect was measured as the ordinal position of the given adjective pair. The pole effect was operationalized as a dummy variable, its value depending on whether the "more similar" scale pole appeared at the left or right. The context effect consisted of two variables. The first reflected the main effect of contextual contamination and was defined to be the average of all ratings preceding the given adjective pair for each individual subject. (It follows that where an adjective pair appeared first no contextual contamination was measureable and these cases were not included in the analyses.) The second variable relating to context effect accommodated its hypothesized interaction with ordinal position. First, the average rating across all adjective pairs and all subjects was subtracted from the "main effect" average described above. This difference, roughly half being negative, was then multiplied by the given adjective pair's ordinal position. This composite variable essentially exaggerates both "high" and "low" contexts as a function of increasing ordinal position. The resulting specified equation, therefore, is

$$\begin{aligned} \text{RATING} = {} & B_0 + B_1\ (\text{ORDINAL POSITION}) \\ & + B_2\ (\text{POLE ORIENTATION}) \\ & + B_3\ (\text{CONTEXT AVERAGE}) \\ & + B_4\ (\text{ORDINAL POSITION})(\text{CONTEXT AVERAGE} - \\ & \text{AVERAGE OF ALL RATINGS}) \end{aligned} \quad (1)$$

for each adjective pair.

The interaction term was dominated by the context effect component. Because of its very high correlation with the main context effect variable, the two are analyzed and interpreted together. The interaction term was not highly correlated with ordinal position (average simple correlation = -.003, average absolute correlation = .113), implying that the latter can safely be considered separately.

Results

Joint Effects

Each adjective pair was analyzed separately, as described above, for both the test and retest phases. Thus, each of the three hypotheses is tested some 20 times. Results are presented in Table 2.

TABLE 2
ADJUSTED COEFFICIENTS OF DETERMINATION, PARTIAL COEFFICIENTS OF DETERMINATION, AND SIGNIFICANCE

		TEST		
Adjective Pair	Adj. R^2	Sequence	Pole	Context
1	-.004	.005	.003	.050
2	.139*	.010	.028	.175*
3	.068*	.001	.011	.128*
4	.410*	.000	.020	.440*
5	.020	.005	.005	.077*
6	.040	.013	.013	.087*
7	.286*	.001	.018	.327*
8	.061*	.001	.027	.096*
9	.244*	.008	.008	.270*
10	.177*	.010	0	.224*

		RETEST		
Adjective Pair	Adj. R^2	Sequence	Pole	Context
1	.136*	.054	.053	.113*
2	.103*	.008	.022	.111*
3	.135*	0	.020	.186*
4	.230*	.017	.017	.263*
5	.110*	.017	.023	.140*
6	.228*	.092	0	.250*
7	.297*	.008	0	.346*
8	.206*	.016	.030	.240*
9	.361*	.004	.088*	.311*
10	.193*	.072	.005	.186*

* = significant at the .10 level

As a descriptive measure of the joint effect of all three types of variables together, adjusted coefficients of determination are presented in Table 2. The average over all 20 instances is .172, indicating a material proportion of variance in ratings can be accounted for by the scale presentation characteristics. Seventeen of the twenty squared correlation coefficients are significant at the .10 level (where two would be expected by chance), indicating that the overall joint effect is significant.

To test for the significance of the three types of effects separately, an F test of the partial correlations of determination (Horton, 1978) was employed.

H1: Sequence Effect

The average value of the partial coefficients of determination over the 20 analyses is .017 and none of the 20 coefficients is significant at the .10 level. It is evident that ordinal position *per se* did not influence subjects' ratings.

H2: Pole Effect

The average value of the partial coefficients of determination over the 20 analyses is .020 and only one of the 20 coefficients is significant at the .10 level (where two would be expected by chance). Clearly, pole orientation of the scales had no material or significant effect on subjects' ratings.

H3: Context Effect

As noted earlier, the context effect consists of two variables, a context main effect measure and an interaction with ordinal position, considered together. The average partial coefficient of determination attributable to these two variables over the 20 instances is .201. Of the 20 coefficients, 19 are significant at the .01 level (where two would be expected to chance). Contrary to the findings of Jain and Pinson (1976), the effect of contextual contamination is very pronounced and significant.

Nature of Contextual Contamination

Of the three sources of bias studied, only the contextual effect was significant. To determine the nature or direction of this influence, it is informative to examine the signs of the simple correlations between ratings and the context "main effect" and "interaction effect" variables defined earlier. All 20 of the rating-main effect and all 20 of the rating-interaction effect simple correlations are positive. The clear implication is that contrary to the findings of Ferber (1952) and Landon (1971), the influence of contextual contamination is direct. The higher the ratings given to earlier stimuli, the higher the ratings given to later stimuli.

Discussion

Taken together, the three potential sources of bias studied had a significant and material effect. Considered separately, no evidence of sequence and pole orientation effects was found while contextual contamination was clearly present.

As expected, particularly in light of the fairly short list of adjective pairs, subjects did not systematically increase or decrease ratings of a given adjective pair as a function of its ordinal position. Also as expected, for want of any substantial theoretical or empirical foundation, subjects did not systematically tend to use either the left or right hand portions of the scale. Ratings of earlier stimuli did indeed influence subsequent ratings although in a direction contrary to that expected.

It is possible that a list of ten stimulus objects is not of great enough length to evoke a bias due to ordinal position. Also, it is possible that randomization mitigates the pole orientation effect although the variable of interest is tendency toward left or right hand portions of the scale, not toward the "more" or "less" extremes.

Within the scope of this research, the consistency of findings across adjective pairs and across test and retest adds to the credibility of the results of this study. That the context effect is so prevalent is all the more noteworthy in that the entire list of adjective pairs was immediately available to subjects. Thus, they were in a position to scan the list, establish extreme stimulus anchoring points, and so on, prior to rating any given adjective pair.

In light of the results of this study, researchers should strongly consider randomizing or otherwise balancing the content of lists of stimulus objects to be rated as to similarity. Or possibly the provision of explicit anchoring stimuli could serve to establish an absolute framework for subjects' ratings.

References

Becker, Sam L., "Why an Order Effect," Public Opinion Quarterly (Fall 1954), 271-8.

Campbell, D. T. and P. J. Mohr, "The Effect of Ordinal Position Upon Responses to Items in a Check List," Journal of Applied Psychology, 23 (February 1950), 62-7.

Day, Ralph, "Position Bias in Paired Product Tests," Journal of Marketing Research, VI (February 1969), 98-100.

Ferber, Robert, "Order Bias in a Mail Survey," Journal of Marketing, 17 (October 1952), 171-8.

Green, Paul E. and Frank J. Carmone, Multidimensional Scaling and Related Techniques in Marketing Analysis (Boston: Allyn and Bacon, Inc., 1970).

Green, Paul E. and Vithala R. Rao, Applied Multidimensional Scaling (New York: Holt, Rinehart and Winston, Inc., 1972).

Greenberg, Allan, "Paired Comparisons vs. Monadic Tests," Journal of Advertising Research, 5 (December 1963), 44-7.

Greenberg, Allan, "Paired Comparisons in Consumer Product Tests," Journal of Marketing, 22 (April 1958), 411-4.

Horton, Raymond L., The General Linear Model (New York: McGraw-Hill, Inc., 1978).

Jain, Arun K. and Christian Pinson, "The Effect of Order of Presentation of Similarity Judgments on Multidimensional Scaling Results: An Empirical Investigation," Journal of Marketing Research, VIII (August 1971), 375-8.

Landon, E. Laird, Jr., "Order Bias, the Ideal Rating, and the Semantic Differential," Journal of Marketing Research, VIII (August 1971), 375-8.

McGuire, W. J., "Cognitive Consistency and Attitude Change," Journal of Normal and Social Psychology, 60 (1960), 345-53.

Myers, James H. and W. Gregory Warner, "Semantic Properties of Selected Evaluation Adjectives," Journal of Marketing Research, V (November 1968), 409-12.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning (Urbana, Ill.: University of Illinois Press, 1957).

Payne, Stanley L., The Art of Asking Questions (Princeton, New Jersey: Princeton University Press, 1951).

Sherif, Carolyn W., Muzafer Sherif, and Roger E. Nebergall, Attitude and Attitude Change (Philadelphia: W. B. Saunders Company, 1965).

Vernon in Bartlett, F. C., et al. (eds.), The Study of Society (London: Kegan Paul Trench Trubner and Co., 1939).