

The Effects of Multiple Specific Items on General Questions: Results From Surveys in Five States

Frederick O. Lorenz, Iowa State University; John Saltiel, Montana State University;
Rob Robertson, University of New Hampshire; Virginia Lesser, Oregon State University
folorenz@iastate.edu, 210 Snedecor Hall, Ames, IA 50011

Key works: question order effects, part-whole questions, experiments in surveys, replication.

Previous research has demonstrated that responses to survey questions are not sensitive to order effects except when two or more questions are perceived to be related. (Smith, 1991). One situation where questions are related is where a general question on a topic either precedes or immediately follows multiple specific items on the same topic. In this article, we examine the effects of multiple specific items on responses to general questions using data from thirteen split-ballot experiments conducted with general populations in Iowa, Montana, New Hampshire, Oregon, and Pennsylvania. In each survey, general questions about overall satisfaction with community life, local government services, or the condition of state highways either preceded (GS) or followed (SG) a series of items addressing specific aspects of the same theme. These experiments were conducted as part of a USDA regional project W-183, "Improvement in rural and agricultural sample survey methods."

Research questions

Three practical concerns have been identified regarding the effects of multiple specific items on responses to general questions. These concerns, expressed as questions, are:

1. Is there evidence that the general question is interpreted as a summary question in the SG order?
2. If the general is placed last, is there evidence or recency effects whereby the last items have a disproportional influence on responses to the general question in the SG order?
3. Do the mean responses to the general question differ in the SG and GS orders and if so, why?

So far, these research questions have been addressed in only three published articles. First, McClendon and O'Brien (1988) and Willits and Saltiel (1995) have found some evidence that when the general question is last, the specific items provide a frame of reference that helps the respondent formulate an answer to the general question. In contrast, when the general question is first, respondents need to recall attitudes and experiences from memory without any clearcut standards. In both articles, the general question was regressed against

responses to the specific items, and the mean square error was smaller and the R-squared was larger in the SG than in the GS order.

Second, both McClendon and O'Brien (1988) and Willits and Saltiel (1995) found evidence of recency effects. When they regressed the general questions against the specific items, they found that the last specific items in the list were stronger predictors of the general question in the SG than GS order.

Third, both Willits and Saltiel (1995) and Schul and Schiff (1993) found evidence that the mean score on the general question was lower when the general question was asked first. Schul and Schiff explained this finding by arguing that when the general question is asked first, respondents search their memory for whatever information they can retrieve, and they are more likely to recall negative than positive experiences.

Methods

Data from eleven experiments in four states are added to one previously published study (Willits & Saltiel 1995). The studies are as follows:

Pennsylvania (Willits and Saltiel, 1995). The general question was, "Compared to life in other areas of the state, how would you rate the OVERALL quality of life in rural areas of Pennsylvania?" Responses ranged from 5 (much better) to 1 (much worse). Respondents were asked to answer the same question on 19 specific domains related to quality of life. (The total sample size was $n=346$).

Montana A: (Willits and Saltiel, 1995) "How would you rate the overall quality of life in your community?" Responses ranged from 1 for poor to 5 for excellent. Nine domains of community life rated on the same scale ($n=997$).

Montana B: "How would you rate the overall quality of life in your community?" Ten aspects of community life were also rated. Eight (8) of these were identical with those used in Montana A. The response scales were also the same ($n=545$).

Montana C: Same as Montana B except that the ordering of the specific items was altered ($n=536$).

New Hampshire A: "How would you rate the overall quality of life in your community?" Responses

ranged from 1 for poor to 5 for excellent. Eleven domains of community were rated using the same scale. Ten of these were identical to those used in Montana B (n = 235).

New Hampshire B. Identical to New Hampshire A except the ordering of the domain items was changed (n = 219).

Iowa A: (Mail and telephone survey) “Please rate the overall quality of services and facilities in (name of town)”. Responses ranged from 4 (very good) to 1 (poor). Respondents also rated nine community services on the same scale (n_m = 2407 and n_t = 889).

Iowa B: (Mail and telephone). “How would you rate the overall quality of government services in (name of town)” Responses range from 4 (very good) to 1 (poor).

Respondents rated seven specific government services on the same scale (n_m = 2778 and n_t = 907).

Oregon A: “Overall, please rate how satisfied you are with the conditions of those highways you have used in the past 12 months”. Responses range from 1 (very dissatisfied) to 5 (very satisfied). Respondents also rated 6 aspects of highways they have traveled in the last 12 months from 4 = excellent to 1 = poor. (n = 1684).

Oregon B and C: Same as Oregon A except that the ordering of the domains differed and the studies were done with different samples (n_B = 849 and n_C = 878).

Results

The results are organized around the three research questions. First, the evidence in Table 1 supports the conclusion that when the general question follows the specific items, respondents interpret the general as a call to provide a summary. When regressing the general against the specific in both orders in all the surveys, the multiple R²s (shown in Table 1) are larger and the mean square errors (not shown) are smaller, for all the regressions in the SG than in the GS order, except one (Montana B). The differences are not noticeably higher for telephone interviews than mail surveys. Overall, there is consistent evidence that, in the SG order, respondents interpret the general question as a call to summarize their responses to the specific items.

Second, the recency hypothesis predicts that the last specific items should exert greater influence than the first items on the general question when the general is last (SG). To look for evidence of recency, we correlated responses to the general question with indices formed by summing the first three (F) and last

Table 1: R-squared for regression of the general question against the specific items in the GS and SG condition.

	<u>GS</u>	<u>SG</u>	<u>Diff</u>
Pennsylvania	0.426	0.522	0.096

Montana A	0.446	0.602	0.156
Montana B	0.477	0.475	-.002
Montana C	0.369	0.548	0.179
New Hampshire A	0.479	0.699	0.220
New Hampshire B	0.534	0.705	0.171
Iowa A (Mail)	0.443	0.465	0.022
Iowa A (Telephone)	0.415	0.475	0.060
Iowa B (Mail)	0.450	0.472	0.022
Iowa B (Telephone)	0.306	0.386	0.080
Oregon A	0.134	0.283	0.149
Oregon B	0.231	0.284	0.053
Oregon C	0.189	0.294	0.111

three (L) items in the list of specifics. In Table 2, the differences between these two correlations (L – F) are shown for the GS and SG condition. The third column presents the difference between the first two columns (SG – GS). The recency hypothesis predicts that this third column should be positive, meaning that the difference between the correlation of the general with the last and with the first (L – F) should be greater in the SG than in the GS order. Of the thirteen experiments, only three (Pennsylvania, Montana A, and Montana B), possibly four (Oregon A), showed evidence of recency effects. The remaining differences except one were close to zero, with Oregon C being that single exception. Overall, we conclude that there is some evidence of recency effects, but it is not strong or consistent.

Table 2: Differences (L –F) between the multiple R of the last three (L) and the first three (F) specific items in the GS and SG orders.

	<u>GS</u>	<u>SG</u>	<u>Diff</u>
Pennsylvania	-.046	0.178	0.224
Montana A	-.111	0.102	0.213
Montana B	-.160	0.016	0.176
Montana C	-.037	-.033	-.004
New Hampshire A	-.022	0.028	0.050
New Hampshire B	0.127	0.085	-.042
Iowa A (Mail)	-.090	-.044	0.046
Iowa A (Telephone)	-.045	-.058	-.013
Iowa B (Mail)	-.146	-.103	0.045
Iowa B (Telephone)	-.094	-.010	0.008
Oregon A	-.082	0.030	0.112
Oregon B	-.037	-.015	0.022
Oregon C	0.168	0.053	-.115

The third research question asks whether respondents’ evaluation of the general question is different if the general precedes rather than follows the specific items. Table 3 presents the mean scores for the general question in both the GS and SG conditions, and the difference between them. Respondent evaluations of the general questions were significantly higher in the SG

condition in eight of the experiments, lower in four, and no different in two (Montana C and Oregon B). The four in which the mean of the general question was significantly lower in the SG than GS order (New Hampshire A and B and Iowa A (mail and telephone)) are particularly interesting because they are contrary to the direction predicted by Schul and Schiff (1993). They argued that the mean of the general question should be lower in the GS than SG order because, in the GS order, respondents will retrieve whatever information they can, and they more likely to recall negative experiences.

Table 3: Mean scores on the general question in the GS and SG conditions.

	<u>GS</u>	<u>SG</u>	<u>Diff</u>
Pennsylvania	5.52	5.74	0.22
Montana A	3.25	3.49	0.24
Montana B	3.40	3.59	0.19
Montana C	3.35	3.33	-.02
New Hampshire A	3.71	3.37	-.34
New Hampshire B	3.67	3.46	-.21
Iowa A (Mail)	2.76	2.56	-.20
Iowa A (Telephone)	2.82	2.70	-.12
Iowa B (Mail)	2.47	2.67	0.20
Iowa B (Telephone)	2.60	2.94	0.34
Oregon A	3.43	3.74	0.31
Oregon B	3.77	3.83	0.12
Oregon C	3.69	3.90	0.21

To probe Schul and Schiff's argument further, we return to their research setting, an evaluation of customer satisfaction with an Israeli public utility (telephone company). They argued that good service is expected of utilities and customers remember unexpected problems. Thus, when the general question is asked first, customers who experienced problems will give lower overall ratings, which reduces the average score. But when the general question is asked last, customer responses will be tempered by recalling both the negative and positive services contained in the list of specific items. This results in a higher average evaluation of the general question in the SG than GS order.

To test whether this argument is consistent with our data, we divided our studies into two groups. The first group included those surveys in which the general question was rated lower in the GS than in the SG order (Table 4(a)). These studies, in turn, can be divided into those that, like Schul and Schiff's (1993) study, asked about the quality of public services (Iowa B and Oregon A, B, and C) and those that asked about community life. The Iowa B studies asked about the quality of local government services and the Oregon studies addressed the quality of highways. The second group included

those surveys in which the general question was rated higher in the GS than SG order (Table 4(b)).

Next, we followed Schul and Schiff's lead by computing a combined score for the three specific items with the lowest evaluations (MIN) and a combined score for the three items with the highest ratings (MAX). Our hypothesis is that if respondents are recalling negative information in the GS order, the correlation between the general question and MIN should be stronger than between the general and MAX in the GS order, but less so in the SG order.

The results in Table 4(a) for the GS order are consistent with Schul and Schiff's prediction for the first 5 studies listed in Table 4(a). For example, in the Iowa B (Mail) survey of satisfaction with government services, the correlation between the general question and the most negatively rated items (MIN) was 0.61, while the correlation between the general and the most positively related items (MAX) was only 0.38. However, this same essential pattern was also found for the SG order, where we expected a much smaller difference. Thus, it seems that the general question correlates more strongly with the most poorly rated specific items regardless of the order. Further, for the Pennsylvania and Montana community surveys, the general questions correlate more strongly with the most positively rated items rather than the negatively rated items in 3 out of 4 of the studies, the only exception being Montana A. Overall, these patterns provide little consistent evidence supporting Schul and Schiff's (1993) hypothesis that the mean scores on the general question are lower in the GS order because, in the absence of any other information, respondents are likely to recall negative experiences. For our surveys, general questions about public services seem to elicit negative recall whereas questions about communities seems to elicit positive recall, but neither has much to do with the explanation of why mean scores are lower in the GS order.

Table 4(b) presents the same correlations for the surveys in which the general question was rated higher in the GS order. Arguing by analogy to Schul and Schiff, we expected the general question to correlate more strongly with MAX than MIN in the GS order, and less so in the SG order. Indeed, the results in the GS order are as we expected; for example, the correlation of the general with MAX is 0.59 but only 0.47 with MIN for New Hampshire A. A similar pattern was found for New Hampshire B, but the patterns were weak for both Iowa B surveys. But again, just as in Table 4(a), we obtain essentially the same pattern for the SG order. Thus, the more positive general rating of communities in the GS order cannot be unambiguously attributed to the

recall of positive, perhaps romantic notions about community.

Table 4: Correlations of the general question with the most positive (MAX) and most negative (MIN) items in the GS and SG order.

(a) Mean score lower in the GS than SG order:

		<u>GS</u>	<u>SG</u>
Iowa B (Mail)	MIN	0.61	0.60
	MAX	0.38	0.43
Iowa B (Telephone)	MIN	0.47	0.57
	MAX	0.35	0.34
Oregon A	MIN	0.34	0.45
	MAX	0.25	0.40
Oregon B	MIN	0.45	0.51
	MAX	0.33	0.30
Oregon C	MIN	0.40	0.52
	MAX	0.33	0.30
Pennsylvania	MIN	0.40	0.34
	MAX	0.43	0.47
Montana A	MIN	0.47	0.50
	MAX	0.44	0.49
Montana B	MIN	0.44	0.49
	MAX	0.54	0.57
Montana C	MIN	0.41	0.44
	MAX	0.50	0.62

(b) Mean score higher in the GS order:

		<u>GS</u>	<u>SG</u>
New Hampshire A	MIN	0.47	0.51
	MAX	0.59	0.62
New Hampshire B	MIN	0.34	0.51
	MAX	0.69	0.72
Iowa A (Mail)	MIN	0.55	0.56
	MAX	0.55	0.53
Iowa A (Telephone)	MIN	0.53	0.56
	MAX	0.57	0.59

Discussion

Our results so far lead to three conclusions for practitioners. First, the results in Table 1 are so consistent that we feel comfortable making recommendations. If a researcher want the general question to serve as a summary of the specific domains, then the general question should come after the specific items, but if a “top-of-the-hat” response is desired, then the general question should come first. There is little evidence that respondents to mail surveys look ahead to specific items in formulating a response to the general question when the general comes first.

Second, although recency effects may be a factor shaping responses to general questions, we doubt that they are major factor. One theme we continue to probe in our research with this data is that recency effects may interact with other characteristics of the specific items, such as the salience or relative abstractness of a domain, to shape its effect on the general question in the SG order, but at this time there is no clear pattern.

Third, it is clear that question order almost always affects the mean of the general question. Only two of the 13 differences reported in Table 3 did not have t-ratios larger than 2.00 in absolute value (Montana C and Oregon B in Table 3). However, it is not clear why. The only hypothesis put forward so far, as expressed by Schul and Schiff (1993), was not consistent with our data. At this point, our recommendation to practitioners is that, if they are interested in comparing general scores over time or across communities or settings, it is prudent to keep the order of the specific items the same and to keep the placement of the general question the same relative to the specific items.

References

- McClendon, M. J. & O'Brien, D. O. 1988. Question order effects on the determinants of subjective well-being. *Public Opinion Quarterly* 52: 351-364.
- Schul, Y. & Schiff, M. 1993. Measuring satisfaction with organizations: Predictions from information accessibility. *Public Opinion Quarterly* 57:536-551.
- Smith, T. W. 1991. Context effects in the general social survey. In P. P. Biemer, et al, (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Willits, F. K. & Saltiel, J. 1995. Question order effects on subjective measures of quality of life: A two-state analysis. *Rural Sociology* 57:654-665.