

A Qualitative Study on the Effects of Grouped versus Interleaved Filter Questions

Aleia Clark-Fobia^{*}, Kathleen Kephart[†], Dawn V. Nelson[‡]

Tags: questionnaire design, interleaf questions, filter questions, cognitive testing

Survey Practice

Vol. 11, Issue 2, 2018

Filter questions are used in many surveys to avoid asking follow-up questions that do not apply to every respondent. Often, detailed follow-up questions are asked immediately after a filter question is endorsed. This design is called interleaving. In contrast, other surveys use a ‘grouped’ approach where several yes/no filter questions are asked first, and the follow-up questions for any filter question that was endorsed are asked later in the survey. Previous research has found that when filter questions are grouped together before follow-up questions, respondents report significantly more “yes” responses than in the traditional interleaved design. Literature also suggests that a decrease in the quality of follow-up responses may be a trade-off of the grouped filter question design. Using a subset of sections from the Consumer Expenditure Survey, we conducted a cognitive study on the effects of grouping versus interleaving filter questions. We present qualitative evidence that the grouped format may increase cognitive burden on respondents and suggest that could be a reason for the observed decrease in response quality to follow-up questions in the grouped filter question design.

INTRODUCTION

Many surveys use filter questions to determine eligibility for detailed follow-up questions. However, evidence is mixed regarding the ideal placement of these follow-up questions. We address two design choices for administering filter and follow-up questions: interleaving and grouping. Interleaving is when the follow-up questions are asked immediately after the relevant filter. Grouping is when follow-up questions are asked after all filter questions are administered.

The potential downside of the interleaved approach is motivated underreporting; studies have found that this question structure leads to fewer affirmative responses to filter questions than the grouping structure. Researchers hypothesize that when respondents learn that answering a filter question negatively will pre-empt a battery of follow-up questions, they answer ‘no’ to shorten the interview (Duan et al. 2007; Eckman et al. 2014; Kessler et al. 1998; Kreuter et al. 2011). In addition to finding evidence of this motivated underreporting in the interleaved design, the (Eckman et al. 2014) verified increased accuracy in the filter questions in the grouped design compared to

* **Institution:** US Census Bureau

† **Institution:** US Census Bureau

‡ **Institution:** US Census Bureau

the interleaved design.

In spite of these findings that favor the grouped design, both (Eckman et al. 2014) and (Kreuter et al. 2011) caution that grouping filter questions comes with trade-offs, one being that recall may be harder for respondents in the grouped format (Eckman et al. 2014). Since respondents are able to remain with one topic at a time in the interleaved format, retrieval may be easier and impose less cognitive burden, leading to improved recall accuracy. (Kreuter et al. 2011) found some support for this hypothesis; they saw significantly more "don't know" or refusal responses to follow-up questions under the grouped structure than the interleaved design. This finding calls into question which design results in less measurement error overall; the grouped design results in higher accuracy of the filter questions, but the interleaved design may result in higher accuracy of the follow-up questions. The authors suggested further research to tease this out.

The Consumer Expenditure Survey (CE), sponsored by the Bureau of Labor Statistics, is a federal survey that provides information on the complete range of consumers' expenditures, incomes, and the characteristics of consumers in the United States. The survey currently uses an interleaved question structure, and thus provided an opportunity to conduct qualitative research on the effects of grouping versus interleaving filter questions.

METHODS

From January to May 2015, six staff members at the Census Bureau's Center for Survey Measurement (CSM) completed 59 cognitive interviews in the Washington D.C. metro area. The protocol included a subset of questions about consumer expenses from the CE survey, grouped into 'sections' of expenses, such as home furnishings, vehicle expenses, and electronics purchase. In total, there were 11 sections in the test (see supplemental materials). Testing took place in the cognitive lab at the U.S. Census Bureau in Suitland, MD, and at offsite locations convenient to participants.

Respondents were recruited through social media advertisements, local organizations, and personal networks. Cognitive interviewers administered the survey protocol to participants followed by a mix of prescribed and spontaneous retrospective probing. Interviews were audio-recorded, and respondents received \$40 for participation.

Potential respondents were screened and required to answer "yes" to at least four screening questions corresponding to the sections included in the interview. Table 1 provides respondent demographics.

Table 1 Respondent demographics.

	Number of respondents
Gender	
Female	39
Male	20
Race	
White	26
Black	31
Asian	1
DK/Refuse	1
Education	
Some college or less	17
College graduate or more	42
Age	
18-50	22
51 and over	37

The CE survey instrument is designed so that respondents are asked filter questions about many potential expenses. After a respondent answers 'yes' to a filter question, they are asked follow-up questions, such as a description of the purchase, the cost, and date purchased. The number of follow-up questions for each of the 11 sections in the testing protocol ranged from 4 to 24. The original CE survey includes an information booklet provided for reference during the interview. The booklet lists section titles and examples of items included as expenses in each section, but it does not include filter or follow-up questions. After several rounds of testing, we added the information booklet to help respondents categorize their expenditures.

The fundamental difference between interleaving and grouping is when follow-up questions are administered. The introductory text and the first filter question were identical between formats, and the difference came when a respondent said "yes" to an interleaved filter question. When a respondent said "yes" in the interleaved format, they were immediately asked a set of follow-up questions about the items reported in the filter question. Upon completing all relevant follow-up questions, they were then asked the subsequent filter questions. In the grouped format, respondents continued on with all filter questions in the section regardless of how they answered and were asked all follow-up questions after the filter questions. The number of filter questions in each section ranged from 5 to 37, so respondents may have answered almost 40 additional filter questions before returning to the follow-up questions for an expense they indicated. In the grouped sections, respondents were reminded what filter question they had said yes to before being asked follow-up questions. Filter questions ranged from specific (e.g., a television) to very broad (e.g., an electronic personal care appliance, such as a hairdryer). The difference between interleaved and grouped question flow is depicted in Figure 1.

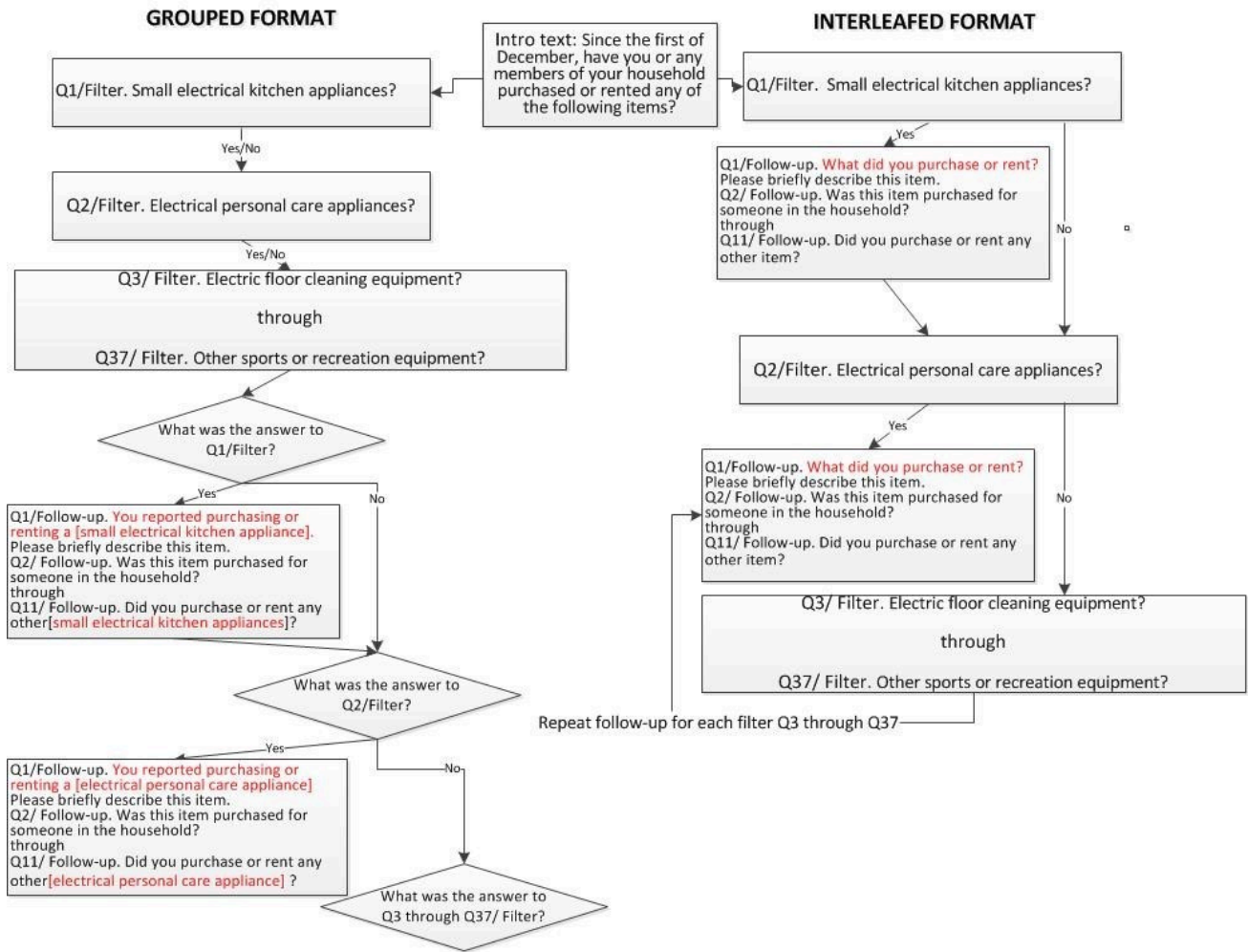


Figure 1 Grouped vs. interleaved question flow.

We used three different protocols to test differences between grouped and interleaved formats. In one protocol, all filter questions were interleaved and in another all filter questions were grouped. In a third protocol, sections of expenses were either grouped or interleaved. We called this a mixed interview since it contained both grouped and interleaved formats. This mixed protocol captures reactions from the same respondent to both formats. In mixed interviews, we varied which sections were grouped and which were interleaved between respondents. The interviews were conducted in three rounds with substantial changes to interviewer training, protocol length, and recruitment criteria between rounds. The distribution of respondents to the protocols is presented in Table 2. There are more interleaved interviews because that format is currently used in the CE, and we were testing other aspects of the interview design independent of the filter question format. We analyzed interview summaries to evaluate whether respondents had difficulty answering questions and what types of problems they commonly encountered. Misreporting, double reporting, and recall difficulty were prominent issues encountered by multiple respondents.

Table 2 Respondents distribution by protocol type.

Protocol Type	Number of Respondents
Grouped	16
Interleaved	29
Mixed- Grouped and Interleaved	15
Total	60[1]

[1] One respondent's data was lost, so only 59 completed interviews are discussed in this report.

RESULTS

The majority of respondents appeared to struggle more to answer the follow-up questions in the grouped format than in the interleaved format, across the mixed and grouped protocols. In the grouped sections, respondents were reminded what filter question they had said yes to before being asked any follow-up questions, for example *"You reported purchasing or renting a [FILL EXPENSE ITEM]. Please briefly describe this item."* Despite this priming, some respondents could not even recall having said yes to the filter question and required additional help from the interviewer and information booklet to aid in remembering the details of the purchase. As one interviewer noted, "The grouping was a disaster; she couldn't remember that she had said yes to some items and sometimes asked for the item number to look at the [information booklet] and help her remember."

Several interviewers observed respondents answering the wrong follow-up questions because they were confused about which filter question they were answering. One interviewer explained, "He would forget what the series of follow-up questions was referring to." Another interviewer also observed this occurring across multiple respondents, "At one point [they] got confused about which item we were on and started to give answers about the previous item again. This issue occurred for at least three of my interviews with a grouped section..."

All five interviewers who conducted mixed interviews saw potential evidence that respondents were expressing more uncertainty and interrupting the interview more to give immediate responses to the follow-up questions in the grouped sections than in the interleaved. While this may have been an artifact of switching between question formats in the mixed protocol, we saw more interruptions with the grouped format, even when the respondent had yet to receive an interleaved section. For example, one interviewer noted that "follow-up was a lot more choppy with grouped than with interleaved and once she [the respondent] clarified what each expense was and what it cost she ended up changing a lot of expenses to 'no' because she had double reported." Another interviewer also noticed an apparent difference in the level of confidence the respondent felt about their answers in the grouped sections of the mixed protocol: "As an interviewer I noticed a distinct difference to how

much more smoothly the interleaved went with her. There was less double reporting or flip flopping; she seemed more confident in when it was purchased and what it cost." Several respondents shared the interviewer's sentiment that the interleaved sections went more smoothly; one respondent described it as follows, "The [grouping] approach of interviewing was harder because it was difficult to follow the flow."

In addition to the interviewer's observations, we probed the 15 respondents who were exposed to both grouped and interleaved questions. Five respondents preferred the grouped format; however, three of those five had issues with double reporting in the grouped format. Ten of the 15 reported strongly preferring the interleaved format. When asked why, respondents said that it was easier to think about the related information (such as the price) at the same time that they were thinking about having purchased it. This would appear to suggest they found it less burdensome to retrieve an inactive memory into working memory, recall all relevant details about that memory, then move on to the next filter question, rather than come back and re-retrieve the memory. This finding supports the theory of recall that (Tourangeau, Rips, and Rasinski 2000) describe in *The Psychology of Survey Response*.

Across multiple respondents we found difficulty with the process of retrieval required by the grouped format. The format requires that they retrieve a memory and then later have to re-retrieve that same memory after recalling an unrelated memory. One respondent who preferred the interleaved design said that he hated going back to an expense when he had moved on to another. He commented, "It's a little less repetitive because you are stopping in the moment when I say 'yes' and asking about it at that moment rather than going through and having to go back and think about why I had answered that question yes."

Other respondents echoed this sentiment and elaborated on how the grouped format affected their response to follow-up questions. Commenting on design preference, one respondent said, "If we go back it's harder because the questions are quite similar, like I wouldn't have mixed up [my purchases]." Another respondent explained, "I liked the immediate follow-up instead of going all the way through, and then going back and saying, 'Okay now you reported...' Because I have that in my train of thought. When you go to the next one I lose it and I have to go back and think again." This process of 'losing it' and having to 'go back and think again' is one potential explanation for the decrease in quality of responses, such as more don't knows or refusals, or changing responses to the filter question.

All interviewers said that it appeared to be easier for respondents to think about an expense and immediately recall and report what it cost and when it was purchased, as opposed to reporting many different expenses and then going back to each expense and reporting details about it later. In the grouped format, many respondents kept interrupting the interviewer to report price and purchase date immediately after answering the filter question. It appeared

more natural for respondents to think about having had an expense and then immediately reporting details of the purchase. These interruptions occurred more in the grouped format even when respondents had not received an interleaved section, so they were not confused about the flow of the interview. This suggests that the information was more readily available to them in that moment.

CONCLUSION

The qualitative data in this study supports findings from the literature on the trade-offs between interleaved and grouped filter questions. Our observations suggest that the grouped format may increase the cognitive burden associated with recall of expenses by requiring respondents to move from topic to topic, rather than remaining with one topic at a time. Respondents exposed to both formats overwhelmingly preferred interleaved and generally expressed that it was easier to recall details when allowed to concentrate on one topic at a time.

The data from our study also suggests that respondents had more difficulty with the follow-up questions in the grouped format. We observed more respondents with grouped filter questions interrupting the interview, expressing doubts about the accuracy of follow-up question responses, and changing their initial answers to the filter questions. This increased difficulty could possibly have affected the quality of the responses to the follow-up questions in the grouped format.

Grouping requires respondents to activate a memory, then stop, and immediately activate a different and potentially unrelated memory. (Tourangeau, Rips, and Rasinski 2000) describe how respondents retrieve a memory from an inactive state in long-term memory and then activate the memory to answer a survey question. In the grouped format, by the time the follow-up questions are asked, a respondent needs to reactivate the initial memory and then attempt to recall details related to the memory. Since memories are generally stored with related events and concepts (Anderson 1983; Collins and Quillian 1969), interleaving may be more conducive to the retrieval of those related details and events that are needed to answer the follow-up questions. Conversely, grouping potentially disrupts this process and could decrease the quality of responses to follow-up questions.

Several limitations exist in our research. Respondents were selected based on having certain expenses, were paid an incentive, and may be more motivated than a typical respondent. Half of our respondents were over the age of 50, and research suggests a negative correlation between age and recall (Herzog et al. 1999). Respondents were also highly educated and may have higher recall ability than average. Also, we have no way to verify if respondents' answers to filters or follow-up questions are accurate and could not measure motivated underreporting.

Further research should validate differences in the quality of data in follow-up

questions between grouping and interleaving. In particular, research should examine if the number of follow-up questions has an impact on data quality that differs between the grouped and interleaved format. Research could also explore balancing the increased cognitive demand of grouped filter questions with the potential for increased endorsement of filter questions.

REFERENCES

- Anderson, J.R. 1983. "A Spreading Activation Theory of Memory." *Journal of Verbal Learning and Verbal Behavior* 22 (3): 261–95. <http://www.sciencedirect.com/science/article/pii/S0022537183902013>.
- Collins, A.M., and M.R. Quillian. 1969. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8 (2): 240–47. <http://www.sciencedirect.com/science/article/pii/S0022537169800691>.
- Duan, N., M. Alegria, G. Canino, T.G. McGuire, and D. Takeuchi. 2007. "Survey Conditioning in Self-reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats." *Health Services Research* 42 (2): 890–907. <http://onlinelibrary.wiley.com/doi/10.1111/j.1475-6773.2006.00618.x/full>.
- Eckman, S., F. Kreuter, A. Kirchner, A. Jäckle, R. Tourangeau, and S. Presser. 2014. "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys." *Public Opinion Quarterly* 78 (3): 721–33. <http://poq.oxfordjournals.org/content/78/3/721.short>.
- Herzog, A.R., W.L. Rodgers, N. Schwarz, D.C. Park, B. Knauper, and S. Sudman. 1999. "Cognitive Performance Measures in Survey Research on Older Adults." In *Cognition, Aging and Self-Reports*, 327–340. Philadelphia: Psychology Press.
- Kessler, R.C., H.U. Wittchen, J.M. Abelson, K. McGonagle, N. Schwarz, K.S. Kendler, B. Knäuper, and S. Zhao. 1998. "Methodological Studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS)." *International Journal of Methods in Psychiatric Research* 7 (1): 33–55. <https://doi.org/10.1002/mpr.33>.
- Kreuter, F., S. McCulloch, S. Presser, and R. Tourangeau. 2011. "The Effects of Asking Filter Questions in Interleaved versus Grouped Format." *Sociological Methods & Research* 40 (1): 88–104. <https://doi.org/10.1177/0049124110392342>.
- Tourangeau, R., L.J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

FIGURES, TABLES, AND SUPPLEMENTARY MATERIALS

Supplemental material 1

Download: <http://app.scholasticahq.com/api/v1/attachments/9620/download>

Supplemental material 2

Download: <http://app.scholasticahq.com/api/v1/attachments/9380/download>
