

# **Effect of Questionnaire Structure on Nonresponse and Measurement Error: Sequential vs. Grouped Placement of Filter Questions**

Lisa R. Carley-Baxter<sup>1</sup>, Andy Peytchev<sup>1</sup>, and Michele Lynberg Black<sup>2</sup>  
<sup>1</sup>RTI International, <sup>2</sup>CDC

## **Abstract**

Based on a cognitive perspective, survey practitioners aim to structure questionnaires in the way that respondents' memories are organized. If details about multiple behaviors are requested, asking the respondent to report a type of behavior and then following up with more detailed questions is assumed to help recall, reduce measurement error and possibly decrease item nonresponse. When surveys ask about multiple behaviors, a fear arises that asking many detailed questions after such a filter question will deter some respondents from reporting the next type of behavior on a filter question, inducing measurement error bias in survey estimates. This may also lead to higher breakoff rates prior to the answering of all key filter questions and higher item nonresponse rates to the follow-up questions. There is some evidence from self-administered surveys that when respondents recognize such a skip pattern, they become more likely to provide responses that avoid subsequent questions. The lack of interviewer presence may encourage this behavior. There is also similar, yet limited evidence on grouping filter questions from in-person interviewing in a study with a relatively long interview. Furthermore, the high interviewer rapport in face-to-face interviews may mask potential effects on breakoff and item nonresponse. We randomly assigned respondents in a dual-frame landline-cell phone study on intimate partner violence to either a sequential or grouped filter question design. In both versions, respondents were asked to identify the number of perpetrators for different types of victimizations. In the sequential design, multiple in-depth follow-up questions about each reported person were asked immediately after each filter question. We evaluate whether key survey estimates, item nonresponse, and breakoff rates differ when multiple in-depth follow-up questions are asked about each reported perpetrator for different types of victimization (sequential) versus asking all filter questions before asking for additional detailed information about the perpetrators (grouped).

## **Introduction**

The structure of the survey instrument can have an impact on survey estimates through multiple types of error. Survey instruments often include follow-up questions that depend on the response to a previous question, or filter question, such as asking about details only if an event or behavior has occurred. To facilitate respondent recall and flow of the questionnaire, the survey instrument can be organized so that each filter question is followed up with related questions on the same topic before moving on to a filter question on another topic. We label this the "sequential" design. As this structure most closely resembles conversational logic and asking all questions on a topic while the respondent is thinking about it prior to moving to the next topic, fewer breakoffs during the interview and less time to administer the survey could be expected. By presenting the questions by topic, recall can also be facilitated

However, there is a potential problem with this structure. In particular, when a large number of follow-up questions are asked for each topic, respondents may alter responses to subsequent filter questions in order to avoid additional questions in order to shorten the length of the survey. This effect may be exacerbated when the survey is on a sensitive topic and when many repetitive questions are asked. Key behaviors measured in a survey may be underestimated because of this aversion to follow-up questions. An alternative is to ask all filter questions together. In an interviewer administered survey, the respondent will not have the ability to learn that their responses to the behavioral (filter) questions lead to additional follow-up questions, removing this reason for underreporting.

Two studies in the research literature suggest that respondents may indeed underreport in order to avoid additional questions. Both Kessler et al. (1998) and Duan et al. (2007) found that respondents reported lower mental distress or mental health service use when a sequential design was used. Both studies, however, were very long face-to-face interviews that may have been particularly taxing to respondents. Many of the follow-up questions were mental health scales with numerous and somewhat similar questions. Furthermore, both studies targeted minority groups that differ from the general population on measures related to cognition and willingness to conduct surveys.

Of interest is to evaluate whether this occurs in a large scale national telephone survey that is, for example, less than an hour in length and is administered without large oversamples of minority groups.

## **Methods**

The target population for the National Intimate Partner and Sexual Violence Surveillance System (NISVSS) is adults aged 18 and older living in households with telephones. The NISVSS is a national random-digit-dial (RDD) survey that collects sensitive information on psychological aggression, coercive control and entrapment, physical violence, stalking, and sexual violence victimization. Sample members were asked to report specific behaviors experienced in their lifetime, the past three years, and the past 12 months. Sample members over the age of 70 were also asked to report past 12 month victimization by a non-intimate partner for psychological aggression, coercive control and entrapment, and physical violence. In addition, respondents were asked to provide their demographic information (including current relationship status), answer some questions about their health, provide detailed information on their perpetrators, and were asked general follow-up questions regarding help and services sought as a result of their victimization. The average survey length was approximately 32 minutes.

The pretest for NISVSS from which we report results used a stratified dual-frame RDD sample design including both landline and cell phone numbers. The landline RDD sample was list-assisted, selected from telephone banks of 100 numbers with at least 1 listed number; known non-working numbers were screened out. The cell phone sample was selected from active telephone banks assigned to cell phone carriers; currently, prescreening cannot be done for cell phone numbers. The landline and cell phone samples were stratified by state to facilitate both

national and state-level estimates in the main study. Sample numbers were purchased from Marketing Systems Group and Survey Sampling International, Inc.

Within each sample, landline and cell phone, cases were randomly assigned to receive multiple in-depth follow-up questions about each reported person immediately after each filter question (sequential design) or to receive the follow-up questions after all filter questions has been administered (grouped design).

For households with only one adult male or one female, that person was selected. For households with both males and females, we randomly selected one sex. In an effort to achieve an approximately equal number of interviews by sex, we oversampled males to compensate for their lower cooperation rate. In households containing two adults, we used a modification of the Rizzo method for selecting a respondent whereby we randomly selected either the person on the phone or the other person in the household. For households with three or more adults, we selected one respondent using the most recent birthday technique.

A six-week pre-test of the NISVSS was conducted in November and December, 2009, including a two-week nonresponse phrase. The purpose of the pretest was to fully test all of the data collection procedures and questionnaire for the full scale implementation in 2010, as well as to conduct an experiment on questionnaire structure regarding placement of filter questions. All cases for which we were able to obtain an address were sent a lead letter introducing the study. A total of 784 interviews were completed; 495 in the sequential structure and 289 in the grouped structure (Table 1). The overall response rate was 12.9% (AAPOR Response Rate #4).

**Table 1. Number of completed interviews by sample type and questionnaire structure.**

Sample Type	Sequential	Grouped
Landline	284	123
Cell phone	97	77
Nonresponse	114	89
Total	495	289

### *Hypotheses*

1. The interview length will be longer in the sequential structure than in the group structure.
2. There will be a higher number of breakoffs, number of behaviors reported, and item nonresponse in the sequential structure than in the grouped structure.
3. Interviewers would prefer the grouped structure over the sequential structure.

### **Results**

*Length of Interview.* No significant differences were found in the length of interview by questionnaire structure assignments. The mean interview length for the sequential structure was 31.4 minutes compared to 31.8 minutes for the grouped structure (Table 2;  $t=-.45$ ,  $p>.7$ ). We also investigated differences by sex, and found no major differences in interview length for either females or males across questionnaire structure.

**Table 2. Length of interview by questionnaire structure and sex of respondent.**

	Females		Males		Total	
	Sequential	Grouped	Sequential	Grouped	Sequential	Grouped
Mean	32.3	32.6	30.3	31.0	31.4	31.8
Median	25.3	25.1	24.3	24.0	24.5	24.6
Min	11.4	11.6	9.1	12.1	9.1	11.7
Max	155.0	144.7	152.4	104.3	115.0	144.7

*Breakoffs.* Table 3 presents the breakoff rate (defined as number of breakoffs / number of cases that started the interview) by questionnaire structure and sex. No significant differences were found between the two questionnaire structures on either overall breakoff rates or breakoff rates by sex.

**Table 3. Overall breakoff rate by questionnaire structure and sex of respondent.**

	Females		Males		Total	
	Sequential	Grouped	Sequential	Grouped	Sequential	Grouped
Breakoff Rate	19.9	23.9	16.7	19.8	18.5	22.0

Next we looked at where in the questionnaire breakoffs happened (before, during, or after victimization questions) to see if any differences could be detected by questionnaire structure or by sex of respondent. The sequential structure had significantly more breakoffs occur within the victimization sections (66.4%) compared to the grouped structure (53.1%). Similarly, for males, significantly more breakoffs occurred within the victimization sections for the sequential structure (70.5%) compared to grouped structure (55.9%). A similar pattern was seen for females, though the difference was not statistically significant.

**Table 4. Percent of breakoffs by section, questionnaire structure and sex of respondent.**

	Females		Males**		Total**	
	Sequential	Grouped	Sequential	Grouped	Sequential	Grouped
Prior to Victimization Sections	24.2	40.4	15.9	44.1	20.9	42.0
During Victimization Sections	63.6	51.1	70.5	55.9	66.4	53.1
After Victimization Sections	12.1	8.5	13.6	0.0	12.7	4.9

\*\*p<.01

*Number of Behaviors Reported.* Similar percentages of behaviors were reported for both sequential and grouped structures (Table 5). No significant difference was found in the mean number of reported behaviors between the sequential (mean=4.99) and grouped (mean=5.08) structures ( $t=-.45, p>.6$ ). Furthermore, no significant differences in the number of behaviors endorsed were found between questionnaire structures for either males and females. For females

the number of behaviors reported was larger for respondents in the sequential (mean=6.34) structure compared to the group structure (mean=5.94) but this was not significant ( $t= 0.43$ ,  $p>.6$ ). For males the sequential (mean=2.78) structure was lower than the grouped structure (mean=3.36); this was borderline significant ( $t= -1.16$ ,  $p>.11$ ).

**Table 5. Number of behaviors reported by questionnaire structure and sex of respondent.**

	Females		Males		Total	
	Sequential	Grouped	Sequential	Grouped	Sequential	Grouped
Mean	6.3	5.9	3.5	4.6	5.0	5.1
Median	2	0	1	1	1	2
Min	0	0	0	0	0	0
Max	47	35	30	47	47	47

*Item Nonresponse.* Nearly 90% of respondents in both structures had no missing data (Table 6). Approximately 11% of respondents in the sequential structure had one to four missing data items compared to 8% of respondents in the grouped structure. Less than 2% of respondents from each group had five or more missing data items. Additionally, no significant difference was found in the mean number of missing data items between the sequential (mean=.51) and grouped (mean=.36) structures ( $t=1.21$ ,  $p>.2$ ). Similarly, there were no substantive differences by sex. For females, the mean number of missing items was .46 in the sequential group compared to .17 for the grouped structure; this difference was marginally significant ( $t= 1.84$ ,  $p>.07$ ). For males, the mean number of missing items was .64 in the sequential group compared to .44 for the grouped structure; this difference was not significant ( $t= .56$ ,  $p>.58$ ).

**Table 6. Item nonresponse by questionnaire structure and sex of respondent.**

	Females		Males		Total	
	Sequential	Grouped	Sequential	Grouped	Sequential	Grouped
Mean	0.5	0.2	0.6	0.4	.5	.4
Median	0	0	0	0	0	0
Min	0	0	0	0	0	0
Max	30	4	34	31	34	31

*Regression Analysis.* To further evaluate the potential impact of the questionnaire structure on key estimates, we ran six Poisson regression equations by sex; five of which used the individual violence types (physical aggression, coercive control and entrapment, physical violence, stalking, and sexual violence) as the dependent variable and the sixth used any reported behavior experienced as the dependent variable. For females, the sequential structure was a significant predictor of increased reporting coercive control and entrapment, physical violence, and sexual violence. However, other predictors, including age, age squared, and lower levels of education (i.e., some high school) had stronger and more consistent effects on reports of experiencing the individual violence types and any behavior experienced (Table 7). As with females, the sequential structure was a significant predictor of decreased reports of coercive control and entrapment and sexual violence for males; in addition it was a significant predictor of stalking (Table 8). Age, age-squared, and some high school education were also stronger and more consistent predictors for reports of experiencing the individual violence types and any behavior for males, similar to the outcome for females.

**Table 7. Poisson regression results for reporting individual violence behaviors and any behavior for females.**

Predictor	Psychological Aggression Perpetrator Count		Coercion & Control Perpetrator Count		Physical Violence Perpetrator Count		Stalking Perpetrator Count		Sexual Violence Perpetrator Count		Any Behaviors Perpetrator Count	
	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)
Intercept	-1.400**	(0.540)	-1.29**	(0.426)	-1.9***	(0.465)	-1.04	(0.586)	-2.20***	(0.509)	-.54	(0.399)
<i>Sequential Order of Filter Questions</i>	0.13	(0.100)	.35***	(0.086)	.21*	(0.089)	-.04	(0.115)	.08*	(0.090)	.06	(0.178)
Age	0.085***	(0.021)	.07***	(0.016)	.08***	(0.017)	.08***	(0.023)	.12***	(0.019)	.07***	(0.014)
Age <sup>2</sup>	0.001***	(.0002)	-.0009***	(0.0002)	-.0009***	(0.0002)	-.001***	(0.0002)	-.001***	(0.0002)	-.001***	(0.0002)
No High School	--	--	--	--	--	--	--	--	--	--	--	--
Some high school	0.7***	(0.204)	.88***	(0.170)	1.43***	(0.184)	.01	(0.259)	.87***	(0.173)	.34*	(0.159)
High school graduate	0.24	(0.172)	.50***	(0.147)	.81***	(0.168)	.11	(0.189)	.15	(0.151)	.10	(0.127)
Technical / vocational	-0.32	(0.291)	.21	(0.213)	.28	(0.246)	.03	(0.280)	-.18	(0.237)	-.15	(0.200)
Some college	0.15	(0.177)	.43**	(0.151)	.47**	(0.176)	-.04	(0.197)	.21	(0.152)	.03	(0.131)
4 year degree	-0.34	(0.208)	-.36*	(0.185)	-.17	(0.209)	.16	(0.203)	-.23	(0.176)	-.08	(0.144)
Non-Hispanic	-0.14	(0.199)	-.04	(0.166)	.20	(0.186)	-.25	(0.215)	.28	(0.205)	-.09	(0.157)
Non-White	-0.06	(0.113)	.15	(0.087)	.06	(0.093)	-.06	(0.128)	-.24*	(0.102)	.00	(0.086)

**Table 8. Poisson regression results for reporting individual violence behaviors and any behavior for males.**

Predictor	Psychological Aggression Perpetrator Count		Coercion & Control Perpetrator Count		Physical Violence Perpetrator Count		Stalking Perpetrator Count		Sexual Violence Perpetrator Count		Any Behaviors Perpetrator Count	
	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)	Coefficient	(s.e.)
Intercept	-3.54***	(0.819)	-.22***	(0.546)	-.26***	(0.678)	.05	(0.694)	-1.01	(0.748)	-0.45	(0.459)
<i>Sequential Order of Filter Questions</i>	-0.161	(0.138)	-.23*	(0.097)	-.19	(0.118)	-.43**	(0.139)	-.60***	(0.144)	-.14	(0.090)
Age	.14***	(0.034)	.11***	(0.022)	.12***	(0.023)	.03	(0.027)	.06*	(0.029)	.05**	(0.178)
Age <sup>2</sup>	-.002**	(0.0004)	-.001***	(0.0002)	-.001***	(0.0003)	-.0004	(0.0003)	-.0007*	(0.0003)	-.0004***	(0.0003)
No High School												
Some high school	.71*	(0.299)	.88***	(0.210)	.10	(0.261)	-.07	(0.319)	.31	(0.331)	.37*	(0.191)
High school graduate	0.16	(0.256)	.40*	(0.183)	.12	(0.198)	-.10	(0.234)	-.09	(0.264)	.03	(0.153)
Technical / vocational	.76*	(0.335)	.60*	(0.262)	.78**	(0.259)	-.12	(0.401)	-.85	(0.616)	.25	(0.233)
Some college	0.422	(0.249)	.45**	(0.183)	.09	(0.201)	-.09	(0.238)	.34	(0.252)	.17	(0.151)
4 year degree	0.223	(0.266)	-.05	(0.201)	-.23	(0.226)	.04	(0.242)	.41	(0.257)	-.06	(0.166)
Non-Hispanic	0.30	(0.265)	.06	(0.165)	-.04	(0.217)	-.56**	(0.224)	-.93***	(0.218)	-.05	(0.165)
Non-White	.43**	(0.157)	.52***	(0.110)	.62***	(0.134)	.07	(0.173)	-.09	(0.182)	.29**	(0.105)

*Interviewer Feedback.* We hypothesized that interviewers would prefer administering all of the filter questions first followed by the detailed questions (grouped structure). Throughout the data collection period we held Quality Circle meetings with the telephone interviewers to provide updates on the data collection, discuss the study progress, and listen to any feedback the interviewers had about the study. We also conducted an interviewer debriefing study near the end of the data collection period. Interviewers consistently reported that they and respondents disliked the grouped structure. The main reason for their dislike was that it was confusing both to the interviewer and to respondents to have to go back and ask follow-up questions about specific perpetrators after the initial reports. They reported that it was easier and clearer for respondents to ask the more detailed questions right after specific perpetrators were reported for a behavior.

## **Discussion**

Very few significant differences were found between the sequential and grouped structures in our bivariate analysis. In particular, no significant differences were found in the length of the questionnaire, breakoff rates, number of behaviors reported, or item nonresponse by structure, either overall or by sex. The only significant difference observed in our bivariate analyses was in where cases broke off; there were significantly more breakoffs within the victimization sections for the sequential structure overall and for men. However, contrary to what we expected, interviewers greatly preferred the grouped structure.

The regression results showed a different story, with the sequential structure being a significant predictor of reporting coercive control and entrapment and sexual violence for both males and females, and also of physical violence for females and stalking for males. What is more surprising, however, is that higher reporting by females was found in the sequential design supporting the flow and recall hypothesis, while higher reporting was found for males in the grouped design supporting the avoidance of additional questions hypothesis. We believe that this set of findings is of critical importance; the interaction between questionnaire structure and gender of the respondent is of theoretical and practical importance.

This interaction with gender suggests that the optimal questionnaire structure depends on the sample members and topics of the questions. Intimate partner violence may be of greater interest to women and we speculate based on these findings that structuring the survey instrument to facilitate topic flow can help reporting. If males find the topic of less interest and are less motivated or engaged to provide accurate responses, the same instrument structure only provides a means to learn how to shorten the survey and an instrument structure that prevents this behavior is preferred.

Overall, our results showed higher breakoffs with the victimization section and higher reports of experiencing specific forms of violence with the sequential structure. These results suggest that while the sequential structure may lead to a greater proportion of breakoffs to occur within a section of detailed behavior questions, overall reports for some of these questions may be higher for the key population group of interest (in this case for females). This could be important for studies where researchers are concerned about potential underreporting of experiences, as is the case with intimate partner and sexual violence victimization.

## References

- Duan, N., M. Alegria, G. Canino, T. McGuire and D. Takeuchi. (2007). Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats. *Health Services Research*, 42, 890-907.
- Kessler, R., H. Wittchen, J. Abelson, K. McGonagle, N. Schwartz, K. Kendler, B. Knauper, 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, 33-55.