Survey Research Methods (2018)
Vol. 12, No. 1, pp. 59-74
doi:10.18148/srm/2018.v12i1.7168

ISSN 1864-3361
http://www.surveymethods.org

Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden

Stephanie Eckman
RTI International
Washington DC, U.S.A.

Frauke Kreuter
Joint Program for Survey Methodology (JPSM)
University of Maryland, U.S.A.
and University of Mannheim, Germany
and Institute for Employment Research (IAB)
Nurenberg, Germany

Looping questions are used to collect data about several similar events, such as employment spells, retirement accounts, or marriages. The loops gather information about the number of events experienced as well as details about each one. The questions require respondents to think hard to recall each event and are often lengthy and repetitive. Looping questions can be asked in two formats, and which format a survey uses may affect the quality of the data collected. We develop hypotheses about the effects of format on measurement error in looping questions and test the hypotheses using experimental data from a web survey with a link to administrative records. Results show that one format collects more accurate event reports, but the other format provides higher quality data to the follow-up questions. We conclude with guidance for those who write survey questions and those who rely on survey data for substantive analyses.

Keywords: survey methodology; burden; measurement error; incentives

1 Background

Surveys often use question loops to collect data about several similar events, such as doctor visits or periods of unemployment. The questions collect information both about the number of such events experienced and details about each one. Looping questions can be asked in two formats, as shown in Table 1. One format, which we call *how-many*, first asks: "In how many different cities have you lived since you were 14?" After giving a number k, the respondent is then sent through k loops of the follow-up questions. There is often no opportunity to change the value k after it has been set. The second format we call *go-again*: "Please tell me about the city where you lived when you were 14 ... How many people lived in that city or town? Did you live in a standalone home, an attached home, or an apartment? ... And in what city did you live after that?" This format requests the same information as the how-many format, but in a different way. It first gathers details about one event before asking if there is another event to report. The crucial difference between the formats is that respondents in the how-many format see the follow-up questions only after they have committed to a value of k, and in the go-again format, respondents see the follow-up questions first and then decide whether to report additional events. As we shall see, this difference in how the formats operate can have complex effects on data quality. This paper develops and tests hypotheses about data quality in the two formats and provides evidence-based guidance to survey researchers who wish to ask looping questions or analyze data collected via such questions.

Both formats are widely used in surveys today. Table 2 gives examples of looping questions in seven surveys in the United States and Europe. This list is not exhaustive of all looping questions in these surveys, but the questions collected here do span a range of topics: personal history, household composition, household finances, and health. In our search through questionnaires of major surveys, we find that the how-many format is more common. Because the formats' implications for data quality have not been thoroughly explored, however, surveys have no good basis for making a choice between the two. We do not see any pattern to when a survey uses the how-many format and when it uses the go-again format.

A conference presentation by Carley-Baxter, Peytchev, and Black (2010) is the only study we are aware of that experimentally compares the how-many and go-again formats of the looping questions. The authors explore the performance of the questions in a telephone survey about sexual violence. The looping questions asked about experiences of psychological aggression, physical violence, stalking and similar behaviors. The how-many format collected more re-

Contact information: Stephanie Eckman, RTI International, 701 13th St NW, Washington D.C. 2005, U.S.A. (Email: seckman@rti.org)

Table 1
Schematic of Two Formats of Looping Questions

How many questions	Go-again questions
How many events X have you experients in your life?	Please think about the first event X that you ever experienced. <follow up questions about event>
Thinking about the first event X . . . <follow up questions about event>	Did you ever experience another event X? <if yes, follow up questions about event>
Thinking about the second event X . . . <follow up questions about event>	Did you ever experience another event X? <if yes, follow up questions about event>
<and so on>	<and so on>

Table 2
Examples of Looping Questions from Current Surveys

Survey	Format	Question text	# loops possible	# follow-up questions
PSID[a]	How many	During <year>, how many full-time or part-time jobs did you/he/she have (not counting work around the house)?	3	8
	Go-again	What is the street address and move-in date of your/head of household's current residence? Have/Has you/he/she lived anywhere else since January?	6	4
HRS[b]	How many	Altogether, how many times have you been married (including your current marriage)?	3	up to 4
PASS[c]	How many	How many friends or family members do you have a close relationship with, not counting members of your HH?	3	5
	Go-again	Was there a cut to the amount of income support you/someone in your household received at any point in time between <start>and <end>? Was there a further cut to the amount of income support during that time?	5	4
NSFG[d]	Go-again	I would like to get some additional information about the people in this household. Is there anyone else who usually lives here?	No limit	5
	How many	How many different females have you ever had intercourse with? This includes any female you had intercourse with, even if it was only once or if you did not know her well.	3	7
ESS[e]	How many	Including yourself, how many people – including children – live here regularly as members of this household?	No limit	3
SHARE[f]	How many	How many children do you have that are still alive? Please count all natural children, fostered, adopted and stepchildren.	20	4
NLSY97[g]	How many	Now I would like to ask you about any college or university experience you've had. How many different colleges or universities have you ever attended?	3	12
	How many	How many of your pregnancies were not live births, that is, they ended in a stillbirth, a miscarriage or an abortion?	No limit	2

[a] Panel Survey on Income Dynamics (Panel Survey on Income Dynamics, 2013)
[b] Health and Retirement Survey (Health and Retirement Survey, 2011)
[c] Panel für Arbeitsmarkt- und Sozialversicherung (Berg et al., 2014)
[d] National Survey of Family Growth (Lepkowski, Mosher, Davis, Groves, & van Hoewyk, 2010)
[e] European Social Survey (Central Co-ordinating Team, 2010)
[f] Survey of Health, Ageing and Retirement in Europe (Börsch-Supan & Jürges, 2005)
[g] National Longitudinal Survey of Youth, 1997 Cohort (Moore, Pedlow, Krishnamurty, & Wolter, 2000)

ports than the go-again format from male respondents (4.6 vs 3.5), but fewer from female respondents (5.9 vs 6.3) (test statistics and significance not reported in slides). The study also finds more missing data in the go-again format but more breakoffs in the how-many format.

Two strains of research about the survey response process provide insight into how respondents answer looping questions. The first is the literature on behavioral frequency questions, which ask respondents to report the number of times an event has occurred, for example: hours of television watched last week or the number of visits to the dentist in the last 3 years. This literature focuses on the burden of the recall task and the strategies respondents use to decrease this burden, such as satisficing. The other relevant body of research concerns motivated misreporting, which argues that respondents manipulate their answers to reduce the length of the interview. This theory is concerned not with the burden of the recall associated with any one question, but with the burden resulting from the length of the questionnaire as a whole. We discuss these theories in more detail below and how each relates to the looping question response process.

Setting aside the follow-up questions for a moment, the how-many question is similar to a behavioral frequency question, which has been the subject of much research (see in particular Blair & Burton, 1987; Burton & Blair, 1991). The optimal response strategy involves comprehending what the question asks, retrieving relevant information, making judgments and estimates based on the retrieved data, and finally reporting the answer (Tourangeau, Rips, & Rasinski, 2000). This process requires careful thought and attention and places a burden on respondents. Respondents who want to shortcut the response process may use rate-based estimation (approximately two 1-week vacations per year over 5 years) or impression-based estimation ("I don't go to the movies that often"). These strategies are prone to errors of over- and underreporting, and respondents using them often gather contextual cues from the question's wording, response options and placement in the questionnaire when formulating their response (Tourangeau & Bradburn, 2010).

Applying these findings to looping questions, we note that the how-many question itself does not provide support to the respondent as she formulates her response. It simply asks for the number of events she has experienced (see Table 1). Because no help is provided, the answer may be of low quality and may be susceptible to satisficing, such as via the rate- and impression-based estimation strategies. The go-again questions, on the other hand, do provide the respondent with some assistance with recall. The behavioral frequency question is decomposed into multiple questions about individual events, often in chronological order. Because the respondent is asked to think only about one event at a time, rather than all events at once, the recall task is likely to be easier and the recall burden lower (Cannell, Oksenberg, Kalton, Bischoping, & Fowler, 1989; Krosnick & Presser, 2010). Additionally, the questions in the go-again format lead the respondent to make several attempts to remember relevant events, and repeated attempts have been shown to improve recall (Schwarz & Oysermann, 2001).

As an example, consider the question from the National Survey of Family Growth in Table 2 about female sexual partners. As a how-many question, it implicitly asks the respondent to recall each partner and count them. The question itself, though, provides no assistance with the recall task: the burden is entirely on the respondent to remember each partner.[1] The go-again version of this question would first ask about one's first female partner, then the second, and so on, breaking the recall task into several smaller steps and structuring the memories in chronological order, which might elicit better reporting.

However, we have so far not considered the role played by the follow-up questions about each event. These questions add to the length of the questionnaire, and longer surveys are more burdensome (Bradburn, 1979; Sharp & Frankel, 1983). The motivated misreporting theory is concerned with this type of burden and the ways that data quality can suffer when respondents try to shorten the interview. Previous studies have found that respondents misreport to filter and screener questions in an effort to reduce the burden of the interview (Tourangeau, Kreuter, & Eckman, 2015), and a similar effect may occur with looping questions. The go-again format makes it clear that reporting an additional event will lead to another loop of follow-up questions. In this format, respondents who want to decrease the length of the questionnaire could underreport the number of events. The how-many format, on the other hand, does not reveal the follow-up questions until the respondent has reported the number of events, hiding the relationship between the number of events reported and the length of the interview, and thus removing the incentive to underreport.

The literature leads us to two competing hypotheses. Based on research into the response process to behavioral frequency questions, we hypothesize that the go-again format leads to more accurate event reports, because it offers more support for the recall process. We call this the *aided recall hypothesis*. The motivated misreporting theory, which holds that respondents take shortcuts to decrease the length of the survey, leads us to hypothesize that the go-again format collects fewer event reports and less accurate event reports. The go-again format makes the relationship between additional events and additional follow-up questions obvious

[1] Once the respondent answers the how-many question and goes through the loop(s), he may remember additional events. In that sense, the how-many *loops* may assist with recall. However, as we discuss below, surveys using the how-many format do not allow respondents to report additional events that they recall after going through the loops.

and thus respondents in this format may underreport events to shorten the survey. We call this the *motivated misreporting hypothesis*.

To test these two hypotheses, we compare the quality of the data collected via the two types of looping questions. We conclude with recommendations for survey researchers who wish to use looping questions and for analysts who work with such data.

2 Experimental Web Survey

To test the quality of the data collected via looping questions, we conducted a web survey that experimentally varied the format of the looping questions. The survey also allowed us to merge responses to administrative records, and thus we can determine not only which format collects more events but also which is more accurate.

The survey and administrative data sources discussed below are not publicly available, because of the sensitivity of the data they contain (living locations, periods of employment and unemployment). The programs we wrote to analyze the data, however, are available at https://github.com/stepheckman/Looping-Questions.

2.1 Sample

The sample of 11,836 named adults was selected from German federal databases (IAB Institut für Arbeitsmarkt- und Berufsforschung, 2011) in three strata representing diverse employment and unemployment histories. The first stratum contained persons who received income support in 2010 and held a social security contributing job in the last 10 years. The second consisted of persons who received unemployment insurance in the last 10 years, held a social security contributing job in the last 10 years, and never received income support. The third stratum consisted of persons who received neither income support nor unemployment insurance and held social security contributing jobs with two or more employers in the last 10 years. Within each stratum, the sample was equal probability.

Each selected person was mailed a letter which asked him or her to go online to complete the web survey. Data collection was from February to April, 2012. The completed sample size was 1,068 cases with a response rate of 9.01% (AAPOR Response Rate 1, American Association for Public Opinion Research (2016)). The median completion time, among those who completed the entire questionnaire, was 20.4 minutes. There were no differences in the mean or median time in the survey or in the looping question sections between the two formats. Paradata indicate that twenty respondents completed the survey on a mobile device. Because of the small size, we do not analyze these respondents separately, but we do include them in our analyses.

An additional 143 cases started the survey but did not finish the entire interview. Those who finished both looping sections ($n = 34$) are included in all of our analyses, for a total of 1,102 cases. See Figure 1 for a flowchart showing the number of cases that reached different points of the survey. Those cases that broke off during the looping sections ($n = 109$) are analyzed only when comparing breakoff rates in the two formats. There were additional cases that broke off before answering any of the looping questions, but we do not consider these cases here at all.

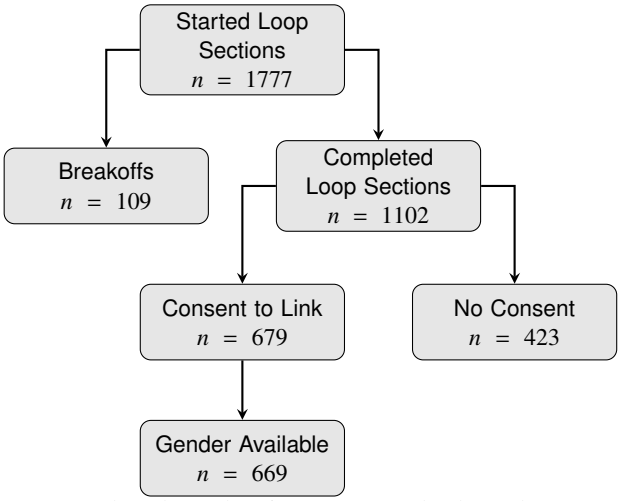

Figure 1. Flowchart Showing Cases' Path Through Survey

2.2 Questionnaire

The survey contained two sections of looping questions. One asked about employers and the other about places the respondent had lived. These questions are similar to those in the PSID (see Table 2), although we asked about lifetime employers and residences rather than those in the last 12 months. We chose these topics because of the expected availability of administrative data with which to evaluate response accuracy (see below).

Each respondent was randomly allocated to receive both looping sections in the how-many format or both in the go-again format. For each mentioned employer or location, the respondent was asked four follow-up questions. The order of the two sections was randomized, with half the respondents asked first about employers and half asked first about locations. The order of the loops within each section was fixed: Respondents were asked to report about their employers/locations in chronological order, starting with the earliest. The full text of all the looping and follow-up questions, translated from German by the authors, is given in the Appendix.

To keep the length of the survey reasonable, and reduce the risk of breakoffs that would harm later experiments, we limited respondents to seven loops through the follow-up questions in the first looping section they saw, and five in the second. Thus, for respondents in the how-many format,

we have the full number of reported employers/locations, but in the go-again format we have at most seven (or five, in the second loop). To compare the responses between the two formats, we re-coded the number of events reported in the how-many format to match what was possible in the go-again format.

After respondents in the how-many condition had reported about all the employers or locations that they had indicated, they were asked if they wanted to report another event.[2] If they said yes, they were sent through the follow-up questions again, and then again asked if they had yet another to report. In our search through other surveys' questions, we do not see this question about additional events asked. We included it to help us understand how respondents answer looping questions.

The web survey we analyze here also contained other manipulations: a consent experiment and an incentive experiment. Both were fully crossed with the assignment to looping question format and the order of the two looping sections. The consent experiment varied the placement and wording of the question about consent to link survey responses to administrative data (see Sakshaug & Kreuter, 2014, for details on the consent questions): we do not expect these manipulations to affect our results (although see Eckman and Haas (2017) for evidence that placing the consent question at the beginning of the survey changes response behavior). The incentive experiment occurred after the looping questions and thus has no effect on our results (Felderer, Kreuter, & Winter, 2013).

2.3 Administrative Data

Because the sample was selected from administrative databases, we are able, with respondent consent, to link the respondents' answers with administrative records. Overall, 61.6% of the respondents ($n = 679$) consented to the link,[3] and the consent rate did not differ by the two looping question formats. The linked administrative records come from the database of social security contributions made each year by employers in Germany (IAB Institut für Arbeitsmarkt- und Berufsforschung, 2013). The database contains reports of all spells of social security contributing employment in Germany since 1975. Each spell contains an identifier for the firm making the contribution and information about the employee as well. From these records we can count the number of different employers each linked respondent has had. The number of employers reported in the administrative data was also re-coded to match the number of possible reports in the go-again condition.

The records captured in the administrative data do not entirely match the common understanding of employment, and thus some mismatch between the survey responses and the administrative data is expected. Positions such as civil servant, police officer, and professor, and self-employment, are not covered by social security and are not captured in the administrative data (Jacobebbinghaus & Seth, 2007). Respondents, however, may include these positions when they report their employers.[4] However, the mismatch between reports and the administrative records should affect reporting in both looping question formats, because of the random assignment of respondents to formats, and thus it should not bias our findings. As a check on the randomization, we tested the difference between the number of employers in the administrative data for those assigned to the how-many format and those assigned to the go-again format, and it was not significant ($F(1, 668) = 2.37, p = 0.12$).

In designing the survey, we had also planned to verify the number of residences, but the concept used in the administrative data (the German *Gemeinde*, an administrative unit that can refer to a city, a portion of a city, or a collection of neighboring villages) was too difficult to ask respondents about. Furthermore, the administrative records on residences only date back to 1995.

3 Methods

The literature review above led us two competing hypotheses: the aided recall hypothesis and the motivated misreporting hypothesis. The first holds that the go-again format leads to better recall, because it reduces the burden of the recall task and takes respondents through their life histories sequentially. The second states that the how-many format collects reports of more events and is more accurate, because it hides the repetitive structure of the loops from the respondents and minimizes underreporting. We designed a web survey specifically to test these two hypotheses about data quality in looping questions.

We consider four aspects of data quality: the number of events reported in the two formats, the accuracy of the number of the events reported, missing data in the follow-up items, and breakoffs. Table 3 summarizes our expectations with regard to each aspect of quality under the two hypotheses developed above.

All analyses are unweighted, because our goal is not to make inference to the population in the three strata used in

[2]This question was asked only if the number of events initially reported was less than the maximum number of trips through the follow-up questions (seven in the first set, five in the second).

[3]The consent rate given here differs slightly from that reported by Sakshaug and Kreuter (2014) for the same survey, because of different decisions about which partially completed cases to include in the analysis data set.

[4]The question text instructed them not to include self-employment, but did not mention the other positions; see Appendix. Given the research showing respondents do not read definitions and other qualifying information in web surveys (Frohlich, 1986; Jenkins & Dillman, 1997), we chose not to make the question longer by describing all the positions to include and exclude.

selection, but to compare the performance of the two formats of looping questions.

3.1 Number of Events

To test how the format effect interacts with other question and respondent characteristics to influence the number of events reported, we run two regression models. The data set for each model is at the section level and thus each respondent appears twice. We account for the correlation between observations for the same respondent with clustered standard errors.

The two models differ only in terms of their case base: the first model includes all respondents and loops (n=1,102, see Figure 1), and the second includes only those respondents who consented to the linkage of administrative data and have gender recorded in the data (n=669). The dependent variable in each model is the number of events reported by the respondent in that loop, re-coded as described above. Because the dependent variable is a count variable, we used Poisson regression models.[5] The independent variables are the format (how-many vs. go-again), the two looping sections (employers vs. locations), and an indicator for the first or second loop. Additionally, the first model includes an indicator of whether the respondent consented to linkage or not, and the second model includes the respondent gender, merged in from administrative data. We include gender because Carley-Baxter et al. (2010) found that the looping questions worked differently for men and women.

3.2 Accuracy of Event Reports

To understand which looping question format leads to more accurate reports of the number of reported events, we use only the response to the number of employers and only those respondents who consented to linkage of their responses to administrative records (n = 679, see Figure 1). For each respondent, we calculate the signed error: the difference between the number of employers reported and the number recorded in the administrative data. We then study the share of correct reports, under- and over-reports by format. The signed error can be negative or positive, but is always a whole number.

3.3 Item Missing Data in Follow-Ups

The quality of reports in the follow-up questions is also of interest to survey data users. Because the administrative data cannot be used to verify the quality of the reports to the follow-up questions in either section, we operationalize data quality in the follow-up items by the fraction of don't knows or refusals.

Both hypotheses lead us to expect that the go-again format will have less missing data in the follow-ups. The aided recall hypothesis suggests that when respondents are in the go-again format, which has them recall events in chronological order and think about them one at a time, they remember details better than if they are in the how-many format and give an estimate for k and then find themselves needing to recall details about each of the k events. The motivated misreporting hypothesis, on the other hand, suggests that respondents in the go-again format simply do not report events that they don't want to answer follow-up questions about; those in the how-many format, who find themselves unexpectedly caught in k loops, are more likely not to answer follow-up questions.

For each loop through the follow-up items, we calculate the number of don't know and refusal responses, out of four, and compare the totals between the two formats. A follow-up counted as missing only when the corresponding event was reported. When a respondent declined to answer one follow-up, the others were still administered.

3.4 Survey Breakoffs

An additional aspect of data quality that we might worry about with looping questions is breakoffs, especially in a web survey. Without an interviewer to encourage continued reporting, web surveys are known to suffer from high breakoff rates (Peytchev, 2009). A respondent who wanted to reduce the burden or length of the questionnaire may simply close her browser window, rather than use the more elaborate underreporting techniques discussed above. Those in the go-again format have another option: they can simply not mention additional events. Thus the motivated misreporting hypotheses leads us to expect more breakoffs in the how may format. The aided recall hypothesis does not have any expectation for the breakoff rates by format.

To test for differences between breakoff rates by format, we use the 109 cases that broke off during the looping sections and analyze the factors that influenced the breakoff decision, using Pearson χ^2 statistics to test for significance.

4 Results

With the web survey data linked to administrative data, we address the four aspects of data quality discussed above and test our two competing hypotheses.

4.1 Number of Events Reported by Format

Figure 2 shows the distribution of the number of events reported, by question and by format. The graphs in the left column refer to the employers section and those in the right column refer to the locations section. The first row shows the number of events reported in the how-many condition, before any re-coding. Those in the second row show the number of

[5]The mean number of events reported is 2.98 with a variance of 3.06. The nearly equal mean and variance support our use of the Poisson model.

Table 3
Expectations for Four Aspects of Data Quality under Aided Recall and Motivated Misreporting Hypotheses

	Hypotheses	
Quality Aspect	Aided Recall	Motivated Misreporting
Number of Events	Go-Again will have more	How-Many will have more
Accuracy of Event Reports	Go-Again will be accurate	How-Many will be accurate Go-Again will underreport
Missing Data in Follow-Ups	Go-Again will have less	Go-Again will have less
Breakoffs	N/A	How-Many will have more

events in the how-many format after re-coding. (Re-coding was to five or seven employers and locations, depending on which loop a respondent was asked first: see Section 2 for more information.) The last row of graphs are for the go-again format, where no re-coding was done, because this format was constrained by the web instrument itself to seven or five loops.

Comparing the first and last rows of graphs in each column, there is clearly a different pattern in the reports between the two formats. Respondents reported more employers and locations in the how-many format than in the go-again format. Even when we re-code the how-many reports to match the maximum number of possible events in the go-again questions, as in the second row of graphs, the number of events reported in the how-many format is still greater than in the go-again format. This result is support for the motivated misreporting hypothesis (see Table 3).

We also see that some respondents reported zero employers, which should not happen given the way we selected the sample. Reporting of zero employers occurred more often in the go-again format (29 times) than in the how-many format (2 times). The higher rate of occurrence in the go-again format is likely the result of underreporting to the filter question that began the employer loop in the go-again format: "Have you ever been employed?" Underreporting to filter questions is a known phenomenon in the motivated misreporting literature (Eckman et al., 2014; Kreuter, Müller, & Trappmann, 2010). It is unfortunate that the inclusion of this filter question introduces a difference in how the two formats worked, but in developing the web instrument we found that the question was needed. Reporting of zero locations in the locations section did not occur and indeed was not possible in either format. There was no filter question at the beginning of the locations loop, because everyone has lived somewhere.

The how-many loops in our questionnaire included a question at the end asking whether respondents had any additional events that they had failed to include when responding to the first question. (See the Appendix for details on how this question was worded.) Ninety-five respondents (22% of those in the how-many condition) reported additional events in one or both sections, leading to 57 additional employers and 68 additional residential locations. One respondent used this technique to report six additional employers and another to report six additional locations – essentially turning the how-many loop into a go-again loop. These additional events are not included in Figure 2, but including them would only strengthen our conclusion that the how-many format collects more events.

The reporting of additional events at the end of the how-many section speaks against the motivated misreporting hypothesis and for the aided recall hypothesis. Recall that respondents in the how-many format first give a count, k, in response to the how-many question itself, then answer k loops of follow-up questions. With this additional question, they then have a chance to report additional events after the k loops. It seems that answering the loops of follow-up questions triggers recall of additional relevant events, as suggested by the aided recall hypothesis. If respondents were interested only in decreasing the length of the survey, as suggested by the motivated misreporting hypothesis, they would say "no" to the question about additional events. In our search through survey questionnaires for looping questions, we saw that surveys do not include the question about additional events at the end of the how-many loops, and thus we exclude these additional events from all analyses. However, we return to this point in the discussion.

Two Poisson regression models, described in Section 3, help us further understand the effects of the two formats on event reports. The dependent variable in each model is the number of events reported in a given loop. Table 4 reports estimated coefficients and marginal effects for both models.

Looking at the first model (columns 1 and 2), in the first row of the table, we see the strong format effect revealed in Figure 2: The go-again format collects one fewer event than the how-many format (marginal effect -0.996). In the second row, we see that respondents report fewer locations than they do employers, which is simply a topic effect. The order of the sections also matters: Respondents report on average

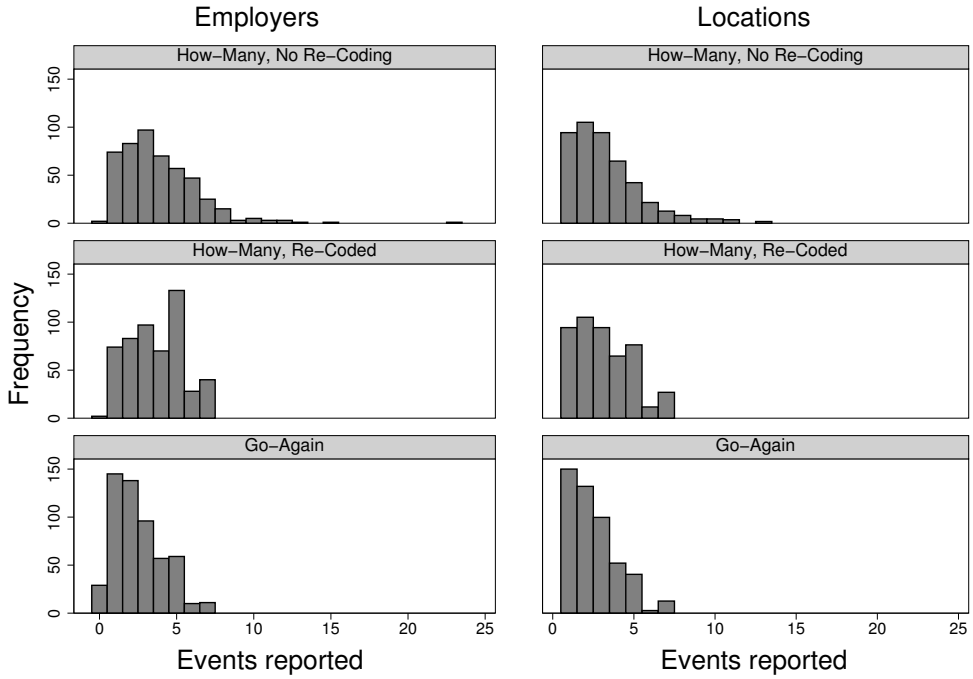

Figure 2. Number of Events Reported, by Format. Additional events reported in the how-many format after the loop(s) completed are not included.

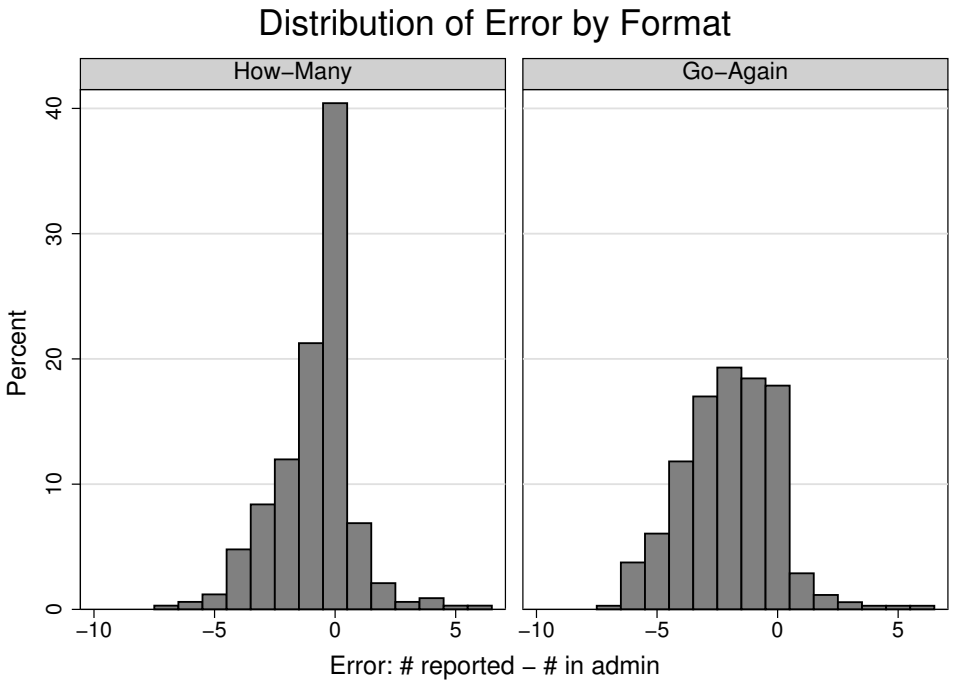

Figure 3. Measurement Error in Number of Employers Reported, by Format. Additional events reported in the how-many format after the loop(s) completed are not included

Table 4
Factors Affecting Number of Events Reported

	Model 1		Model 2	
Variable	$\hat{\beta}$ (Std. Err.)	Marg. Effect (Std. Err.)	$\hat{\beta}$ (Std. Err.)	Marg. Effect (Std. Err.))
Format *(reference category: How-Many)*				
Go-Again	−0.478*	−0.996*	−0.302*	−0.831*
	(0.0506)	(0.0766)	(0.0541)	(0.0975)
Section *(reference category: Employers)*				
Locations	−0.0916*	−0.273*	−0.0997*	−0.303*
	(0.0214)	(0.0639)	(0.0277)	(0.0841)
Section Order *(reference category: First)*				
Second	−0.142*	−0.339*	−0.139*	−0.328*
	(0.0269)	(0.0636)	(0.0356)	(0.0839)
Consent to Link *(reference category: No)*				
Yes	−0.0145	0.163*		
	(0.0348)	(0.0795)		
Gender *(reference category: Female)*				
Male			−0.000634	−0.0183
			(0.0428)	(0.0972)
Interactions				
Go-Again × Consent	0.170∗			
	(0.0555)			
Go-Again × Second Section	0.0683		0.0702	
	(0.0434)		(0.0555)	
Go-Again × Male			−0.0122	
			(0.0645)	
n Loops	2,204		1,338	
n Respondents	1,102		669[a]	
F test of model	$F(6, 1096) = 35.91^*$		$F(6, 663) = 16.47^*$	

Note: Dependent variable is number of events reported in looping section. Dependent variable does not include additional reports in how-many format. Standard errors adjusted for the fact that each respondent appears twice. Constant not shown
[a] Although 679 cases consented to linkage, 10 were missing data on gender
* $p < 0.05$

one-third fewer events in the second section of looping questions than in the first. There are two possible explanations for this finding. Respondents were not able to report more than seven in the first set and five in the second set, and for this reason, we see reduced reporting in the second section. It is also possible that respondents learn and remember across sections and report fewer events in the second section. Unfortunately, our design does not let us distinguish between these two possible explanations. Those who provide consent to merge in administrative data report slightly more events (marginal effect 0.163). The format effect is slightly weaker for consenters than nonconsenters: we suspect that those who consent are generally more cooperative respondents who are less likely to engage in motivated underreporting, but no research has explored this connection. The format effect is not stronger or weaker in the second section.

The third and fourth columns show the results from a similar model run on a subset of respondents, those who consented to administrative record linkage. The results are substantively the same, which supports our use of the cases that consented to linkage to address the issue of response accuracy below. We do not see a difference in reporting behavior for men versus women, nor a different format effect for men, in contrast to Carley-Baxter et al. (2010): Their finding that

the two formats worked differently for men and women may have been the result of the sensitive nature of the topic of their survey, violence in sexual relationships.

4.2 Accuracy of Event Reports

The linked administrative data lets us examine measurement error in the number of employers reported. As discussed in the Section 2, we do not expect the administrative data to match the survey reports in all cases, because of mismatches between respondents' understanding of employers and what is captured in the administrative data, which are derived from records of social security contributions. However, because respondents were randomly allocated to the two looping question formats, we should not see differences in the reporting patterns between the formats unless they perform differently.

Figure 3 shows the signed error in both formats. We see many more correct reports in the how-many format. In both formats, underreports (error < 0) are more common than overreports, but there are many more underreports in the go-again format. These results support the motivated misreporting hypothesis and are in line with our findings with filter questions (Eckman et al., 2014). When collecting the correct number of events is important in a survey, the how-many format performs better than the go-again format.

4.3 Item Missing Data in Follow-Ups

The quality of responses to the follow-up questions are also of interest to data analysts. Because we cannot validate responses to the follow-up questions with administrative data, we instead compare the missing data rates between the formats.

Figure 4 shows the average number of times a respondent answered "don't know" or refused to answer a follow-up item, by each event reported. The vertical axis is the event number: respondents could report up to seven events in the first loop and up to five in the second loop, although, as we saw in Figure 2, many reported fewer.

The horizontal axis is the average number of missing responses to the four follow-up questions. This analysis combines the employer and location sections, so that each respondent contributes up to two data points to the graph. The two sections perform similarly.

There are strong differences in the quality of follow-up information collected in the formats. Respondents in the how-many format were much more likely to say "don't know" or to refuse to answer the follow-up questions, than those in the go-again format. Even when reporting about the first event, respondents in the how-many format do not answer 0.6 of the four follow-up questions, on average. For those in the go-again format, slightly fewer than 0.5 of the four questions have missing values. At the fifth event, the quality of the responses to the how-many format has decreased: Respondents do not provide answers to 1.9 of the four follow-up items, on average. In contrast, the reporting of those in the go-again format has improved and respondents fail to answer fewer than 0.3 follow-up questions. The difference between the two formats is statistically significant, as shown in Figure 4 by the non-overlapping confidence intervals. In both sections, the start and end dates of each spell were most often missing.

Both sections asked about the earliest event first, and the follow-up questions about this event might be the hardest for respondents to answer, because of memory decay (Tourangeau et al., 2000, Section 3.3.1). As the loop continues, respondents are reporting about more recent employers or locations, which should make recall easier, and thus we might expect fewer missing values to later events in both formats. We do see less missing data with later events in the go-again format: the point estimates in Figure 4 shift to the left as we move down the vertical axis and forward in time. In the how-many format, however, the quality of the responses to the follow-up questions deteriorates (shifts right) as we ask about later events. By the seventh job or location, these respondents are answering about a relatively recent event, but are failing to answer 2.6 out of the four follow-up questions.

These results are consistent with both of our hypotheses. It may be that lack of motivation to provide high-quality data drives the missing data rates in the follow-up questions in the how-many format. Other interpretations are also possible, however. It could be that respondents in the go-again format choose not to report an event if they do not know the answers to the follow-up questions.

4.4 Breakoffs

Nine percent (n=109) of the cases that started the first looping section exited the web survey before finishing both looping sections (Table 5). Breakoffs were slightly more common in the how-many format (10.4%) than in the go-again format (7.6%), but the difference is not significant. Breakoffs were more likely when the employer section was asked first (11.0%) than when the locations section was first (7.0%). These results are not in line with either hypothesis.

5 Discussion and Conclusion

The how-many and go-again formats collect different numbers of events, and the higher reports in the how-many format appear to be more accurate. However, there is a trade-off between the number of events reported and the quality of the data collected by the follow-up items: Respondents in the how-many format are more likely not to answer the follow-up items. Although the breakoff rate is high, about ten percent, there is no significant difference in the rates between the two formats.

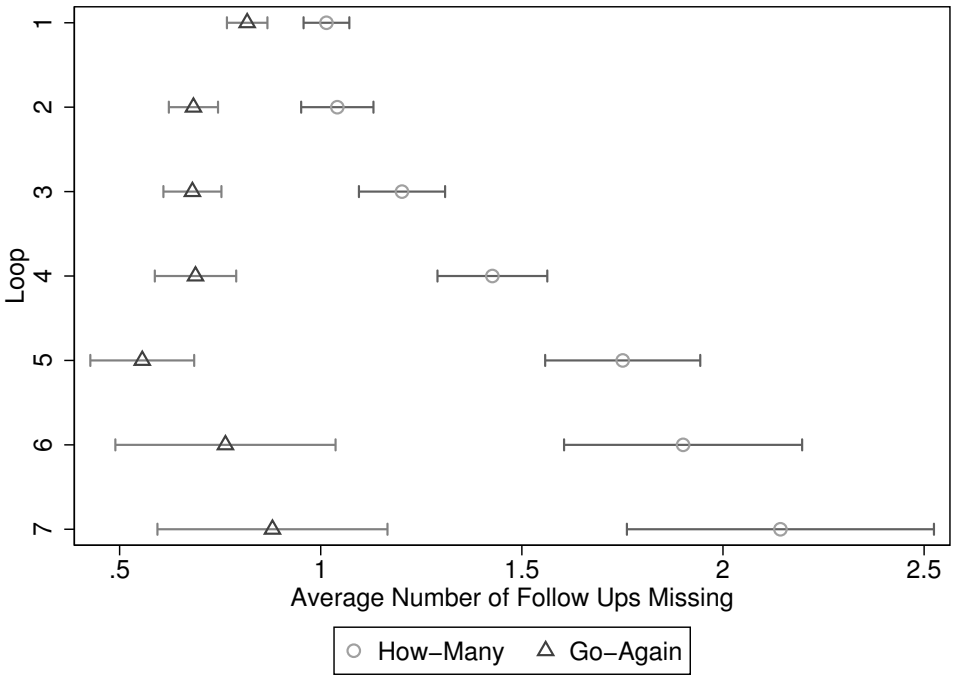

Figure 4. Average Number of Missing Follow Ups, by Event. Reported & Format. Standard errors adjusted for the fact that each respondent may appear more than once.

Table 5
Breakoff Cases, By Format and Section

	n	Breakoffs (%)	Test Statistic[a]
Overall	109	9.0	
Format			
How-Many	64	10.4	2.89
Go-Again	45	7.6	
First Loop			
Employers	67	11.0	6.08*
Locations	42	7.0	

[a] $\chi^2(1)$ test of difference in breakoff rates
* Difference significant at 5% level

The results support the motivated misreporting hypothesis: When the connection between the number of events reported and the number of follow-up questions asked is made explicit, as it is in the go-again format, respondents report fewer events. Although the go-again format may aid respondents' recall of events, any positive effects of the assistance with recall seem to be overshadowed by the motivation to answer fewer loops of repetitive follow-up questions.

We do find some support for the aided recall hypothesis, however. In our version of the how-many format, respondents were given an opportunity to report additional events after the last loop in a section. If in fact the loops had aided recall and led the respondent to recall one or more additional events, those events were captured. We were surprised to find that 22% of the respondents in the how-many format reported one or more additional events to this question. This result suggests that answering the follow-up items in the how-many loops did stimulate recall and that some respondents did not endeavor to skip additional loops of follow-up questions. Nevertheless, on balance the evidence is in favor of motivated misreporting in looping questions.

This study has a few limitations. The first is the use of a filter at the beginning of the go-again loop in the employers section, which introduces another point at which respondents can shorten the questionnaire. Because the go-again version began with a filter question, it was already at a disadvantage relative to the how-many format, though only in the employers section. A cleaner experimental comparison of the two looping formats would not use a filter question. We encourage future research on looping questions to avoid this question when possible. Second, this research was done in only in the web mode and only in Germany. Previous studies of motivated misreporting have shown that the broader phenomenon occurs in all countries and modes tested (Tourangeau et al., 2015). The third limitation is the fact that both of our loops asked questions about events in chronological order. We did not test a backwards ordering, as some research suggests may aid recall (Loftus & Fathi, 1985; Schwarz & Oysermann, 2001; Whitten & Leonard, 1981).

Future work should experiment with the order of the loops to see how responses differ. Furthermore, this study asked looping questions on only two topics, employers and living locations. Both types of events are at the top level of the memory hierarchy (Belli, 1998), and thus recall is unlikely to be a problem for most respondents. With other topics, for which respondents may not as easily know the true answer, the how-many format question may elicit more satisficing, that is more, rate- or impression-based estimation. In such cases, the aided recall hypothesis may hold, and future research should explore this issue.

An alternative to looping questions for collecting life history data is the event history calendar, a form of less standardized interviewing which involves asking respondents about big life events to provide structure and aid in recall. For example, to collect an employment history, a survey may start by having the respondent fill in a calendar of moves, job changes, and relationships. Such events are at the top level in the memory hierarchy and thus help give structure to the recall task (Belli, 1998). Only afterward are respondents asked follow-up questions about each job, such as occupation or pay, Belli, Shay, and Stafford (2001) and Belli, Smith, Andreski, and Agrawal (2007) compare the event history calendar to questionnaires that use a mixture of filter questions and go-again and how-many loops. They find that the calendar usually collects more events and is more accurate. The calendar seems to provide more cues and context that aid the recall task (Belli, Lee, Stafford, & Chou, 2004). It also facilitates the cleaning up of spells to ensure that they are complete and consistent.

Event history calendars are a promising approach to the collection of life history data, and Schwarz and Oysermann (2001) recommend their use in other contexts as well, such as short-period recall. However, it is clear from Table 2 that surveys have not discontinued the use of looping questions, and thus additional research into the best way to ask such questions is important. Our results demonstrate that respondents are sensitive to interview length and repetitiveness. Slowing the interview down to assist with recall, as the event calendar approach does, may backfire and cause respondents to underreport. We encourage further research to compare the two approaches on studies of the general population: the comparisons conducted to date (Belli et al., 2001; Belli et al., 2007) have been with long-time respondents to the Panel Survey on Income Dynamics, who may be less prone to motivated misreporting than survey respondents generally.

We also suggest further study with a hybrid type of looping question that asks about the different events one at a time ("Where did you live when you were 14?... And where did you live after that?... And where did you live after that?") but then asks all of the follow-up questions at the end. We have found a few instances of such looping questions in the questionnaires we searched through. This approach may provide the benefits of assisted recall, by asking several times about related events in chronological order, without being susceptible to motivated misreporting, because the follow-up questions are not revealed until the respondent has committed to a number of events.

The finding of motivated misreporting in looping questions, together with previous evidence of similar misreporting in filter and screener questions (Tourangeau, Kreuter, & Eckman, 2012, 2015), underscores the importance of incentivizing respondents to provide high-quality data. It is not enough to write clearly worded questions and to persuade respondents to participate in the survey – evidence is mounting that data quality can suffer if respondents find the questions repetitive or burdensome, and survey designers need to acknowledge and adapt to this phenomenon.

5.1 Practical Advice

Because looping questions are widely used in surveys, and repetition in survey questions cannot be entirely avoided, we close with practical advice for survey researchers. In choosing how to ask looping questions, survey designers need to consider which data are of highest importance to later analyses. When the number of events is most important, the how-many format should be used. However, this choice may result in a high rate of missing data in the follow-ups. In other research contexts, the goal may be to collect detailed information only about a few events: Table 2 shows that many surveys cap the number of loops of follow-up questions that are asked. In these cases, the go-again format may be a better choice. However, researchers analyzing data collected via the go-again format should keep in mind that the number of events reported is likely too low.

The how-many format does have one important advantage over the go-again format. In the how-many format, the missing data in the follow-ups is obvious, and imputation could be used to fill in the data, making a complete data set. In the go-again format, however, it is not clear that entire events are missing, and analysts can easily overlook this fact. Moreover, because it is not known how-many events are missing for each respondent, imputation is less useful. Although the data set provided by the go-again format appears more complete, because there are fewer cells with missing value codes, the number of events collected by the how-many format is more accurate, and thus the responses may contain more useful information for analysts. Researchers may favor the how-many format for this reason.

We strongly recommend that researchers using the how-many format include an additional question at the end asking if there are any other events the respondents wants to mention. This question gives respondents a chance to report any events that they recalled while answering the follow-up loops, and in this way takes advantage of all of the available recall cues.

6 Acknowledgments

Funding for this project was provided by the Institute for Employment Research (IAB) and by Ludwig Maximilian University. We are grateful for assistance given by Barbara Felderer, Annette Jäckle, Antje Kirchner, Yuliya Kosyakova, and Ulrich Thomsen. Comments by Ruben Bach, Bob Belli, Georg-Christoph Haas, Arie Kapteyn, MaryBeth Ofstedal, Norbert Schwarz, Roger Tourangeau and journal reviewers improved the paper. Daniel Fehrle, Hannah Mautner, and Svenja Wippich provided research assistance.

References

American Association for Public Opinion Research. (2016). *Standard definitions: final dispositions of case codes and outcome rates for surveys*. AAPOR. Lenexa, Kansas.

Belli, R. F. (1998). The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory*, *6*(4), 383–406. doi:10 . 1080 / 741942610

Belli, R. F., Lee, E. H., Stafford, F. P., & Chou, C.-H. (2004). Calendar and Question-List Survey Methods: Association Between Interviewer Behaviors and Data Quality. *Journal of Official Statistics*, *20*, 185–218.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly*, *65*(1), 45–74. doi:10.1086/320037. eprint: http://poq.oxfordjournals.org/content/65/1/45.full.pdf+html

Belli, R. F., Smith, L. M., Andreski, P. M., & Agrawal, S. (2007). Methodological Comparisons Between CATI Event History Calendar and Standardized Conventional Questionnaire Instruments. *Public Opinion Quarterly*, *71*(4), 603–622. doi:10.1093/poq/nfm045. eprint: http://poq.oxfordjournals.org/content/71/4/603.full.pdf+html

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., … Trappmann, M. (2014). *Codebuch und Dokumentation des "Panel Arbeitsmarkt und soziale Sicherung" (PASS) Band III: Personendatensatz (PENDAT)*. Retrieved from http://doku.iab.de/fdz/reporte/2014/DR%5C_02-14%5C_III.pdf

Blair, E. & Burton, S. (1987). Cognitive Processes Used by Survey Respondent to Answer Behavioral Frequency Questions. *Journal of Consumer Research*, *14*, 280–288.

Börsch-Supan, A. & Jürges, H. (2005). *The Survey of Health, Ageing and Retirement in Europe – Methodology*.

Bradburn, N. M. (1979). Respondent Burden. In *Health survey research methods: second biennial conference, williamsburg, va*.

Burton, S. & Blair, E. (1991). Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys. *Public Opinion Quarterly*, *55*, 50–79.

Cannell, C., Oksenberg, L., Kalton, G., Bischoping, K., & Fowler, F. J. (1989). *New Techniques for Pretesting Survey Questions*. Survey Research Center University of Michigan.

Carley-Baxter, L. R., Peytchev, A., & Black, M. L. (2010). Effect of Questionnaire Structure on Nonresponse and Measurement Error: Sequential vs. Grouped Placement of Filter Questions. Phoenix, AZ. Presentation at American Association for Public Opinion Research Conference.

Central Co-ordinating Team. (2010). *European Social Survey Round 3 2008/2009* (Final Activity Report No. ESS4e03.0). City University London.

Eckman, S. & Haas, G.-C. (2017). Does granting linkage consent in the beginning of the questionnaire affect data quality? *Journal of Survey Statistics and Methodology*. doi:doi.org/10.1093/jssam/smx016

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Presser, S., & Tourangeau, R. (2014). Assessing the Mechanisms of Misreporting to Filter Questions. *Public Opinion Quarterly*, *78*(3), 721–733.

Felderer, B., Kreuter, F., & Winter, J. (2013). Can We Buy Good Answers? The Influence of Respondent Incentives on Item Nonresponse and Measurement Error in a Web Survey. Presented at the American Association for Public Opinion Research Annual Conference, Boston, MA.

Frohlich, D. (1986). On the organization of form-filling behaviour. *Information Design Journal*, *5*(1), 43–59.

Health and Retirement Survey. (2011). *Sample Sizes and Response Rates*. http://hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf.

IAB Institut für Arbeitsmarkt- und Berufsforschung. (2011). *Nuremberg: Integrierte Erwerbsbiographien (IEB) V09.00*.

IAB Institut für Arbeitsmarkt- und Berufsforschung. (2013). *Nuremberg: Integrierte Erwerbsbiographien (IEB) V09.03*.

Jacobebbinghaus, P. & Seth, S. (2007). The German Integrated Employment Biographies Sample IEBS. *Schmollers Jahrbuch Zeitschrift für Wirtschafts- und Sozialwissenschaften*, *127*, 335–342.

Jenkins, C. R. & Dillman, D. A. (1997). Towards a Theory of Self-Administered Questionnaire Design, in Survey Measurement and Process Quality. In L. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N.

Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 165–197). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118490013.ch7

Kreuter, F., Müller, G., & Trappmann, M. (2010). Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. *Public Opinion Quarterly*, *74*(5), 880–906. doi:10.1093/poq/nfq060. eprint: http://poq.oxfordjournals.org/content/74/5/880.full.pdf+html

Krosnick, J. A. & Presser, S. (2010). Question and Questionnaire Design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 263–313). Emerald Group Publishing Limited.

Lepkowski, J. M., Mosher, W. D., Davis, K., Groves, R. M., & van Hoewyk, J. (2010). The 2006–2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. *Vital Health Statistics*, *2*(150).

Loftus, E. F. & Fathi, D. C. (1985). Retrieving Multiple Autobiographical Memories. *Social Cognition*, *3*(3), 280–295.

Moore, W., Pedlow, S., Krishnamurty, P., & Wolter, K. (2000). *National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report*.

Panel Survey on Income Dynamics. (2013). *PSID Main Interview User Manual*. http://psidonline.isr.umich.edu/data/Documentation/UserGuide2011.pdf.

Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly*, *73*(1), 74–97. doi:10.1093/poq/nfp014. eprint: http://poq.oxfordjournals.org/content/73/1/74.full.pdf+html

Sakshaug, J. W. & Kreuter, F. (2014). The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey. *Public Opinion Quarterly*, *78*(1), 166–176. doi:10.1093/poq/nfu001. eprint: http://poq.oxfordjournals.org/content/78/1/166.full.pdf+html

Schwarz, N. & Oysermann, D. (2001). Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation*, *22*(2), 127–160.

Sharp, L. M. & Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly*, *47*(1), 36–53.

Tourangeau, R. & Bradburn, N. M. (2010). The Psychology of Survey Response. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 315–346). Emerald Group Publishing Limited.

Tourangeau, R., Kreuter, F., & Eckman, S. (2012). Motivated underreporting in screening interviews. *Public Opinion Quarterly*, nfs033.

Tourangeau, R., Kreuter, F., & Eckman, S. (2015). Motivated Misreporting: Shaping Answers to Reduce Survey Burden. In U. Engel (Ed.), *Survey measurements. techniques, data quality and sources of error* (pp. 24–41). Frankfurt/New York: Campus.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge. Cambridge University Press.

Whitten, W. B. & Leonard, J. M. (1981). Directed Search Through Autobiographical Memory. *Memory & Cognition*, *9*(6), 566–579.

Appendix
Text of Looping Questions

Employers, How-Many Format

- How many employers have you worked for in your life so far? Please do not count periods when you were self-employed.
- LOOP THROUGH FOR EACH REPORTED EMPLOYER UP TO 7 (in first section) or 5 (in second section)
  - Please think about your first/second/. . . employer. When did your employment there begin? (Month, Year)
  - When did your employment there end? (Month, Year or Ongoing)
  - How many hours did you work per week in that job? (Hours)
  - What was your position when you stopped working there? (5 Categories: Blue-collar Worker, White-collar Worker, Member of Military, Civil Servant or Judge, Employed in Family Business)
- IF FEWER THAN 7 (5) EMPLOYERS REPORTED: Did you have any other employers that you didn't include before?
- IF YES:
  - Please think about your first/second/. . . employer. When did your employment there begin? (Month, Year)
  - When did your employment there end? (Month, Year or Ongoing)
  - How many hours did you work per week in that job? (Hours)
  - What was your position when you stopped working there? (5 Categories: Blue-collar Worker, White-collar Worker, Member of Military, Civil Servant or Judge, Employed in Family Business)
- IF FEWER THAN 7 (5) EMPLOYERS REPORTED: Did you have any other employers that you didn't include before? <continued until respondent said no, or reported about maximum number of events >

Employers, Go-Again Format

- Have you ever been employed? Please do not count periods when you were self-employed.
- IF YES:
  - Please think about your first/second/. . . employer. When did your employment there begin? (Month, Year)
  - When did your employment there end? (Month, Year or Ongoing)
  - How many hours did you work per week in that job? (Hours)
  - What was your position when you stopped working there? (5 Categories: Blue-collar Worker, White-collar Worker, Member of Military, Civil Servant or Judge, Employed in Family Business)
- Did you have any other employers after that?
  - IF YES, LOOP THROUGH FOLLOW UP ITEMS AGAIN
- <continued until respondent said no, or reported about maximum number of events >

Residential Locations, How-Many Format

- How many places [cities/towns] have you lived in?
- LOOP THROUGH FOR EACH REPORTED LOCATION UP TO 7 (in first section) or 5 (in second section)
  - Please think about your first/second/. . . place of residence. When did you begin living there? (Month, Year or since I was born)
  - And when did you move away? (Month, year, have not moved away)
  - How many people lived there? (Number)
  - Which state is that in? (Choose from 16 federal states)
- IF FEWER THAN 7 (5) LOCATIONS REPORTED: Have lived anywhere else that you did not yet mention?
- IF YES:
  - When did you begin living there? (Month, Year or since I was born)
  - And when did you move away? (Month, year, have not moved away)
  - How many people lived there? (Number)
  - Which state is that in? (Choose from 16 federal states)
- IF FEWER THAN 7 (5) LOCATIONS REPORTED: Have lived anywhere else that you did not yet mention? <continued until respondent said no, or reported about maximum number of events >

Residential Locations, Go-Again Format

- Please think about your first/second/. . . place you have lived.
  - When did you begin living there? (Month, Year or since I was born)
  - And when did you move away? (Month, year, have not moved away)
  - How many people lived there? (Number)
  - Which state is that in? (Choose from 16 federal states)
- Have you lived anywhere else?
  - IF YES, LOOP THROUGH FOLLOW UP ITEMS AGAIN
- <continued until respondent said no, or reported about maximum number of events >