Testing New Interview Protocols: Lessons Learned About Interviewers, Respondents, and Survey Content

Erica Yu[1], Jennifer Edgar[1], Brandon Kopp[1], Brett McBride[1]

[1]Bureau of Labor Statistics, 2 Massachusetts Avenue, NE, Washington, DC 20212

Abstract

The Gemini redesign proposal for the Consumer Expenditure Survey includes an emphasis on records to aid in the reporting of difficult-to-recall expenditures. Research was needed to understand how this new interview protocol would affect the length, burden, and respondent use of records during the interview. For this study, two new protocols were tested: 27 participants were assigned to a "respondent-track" protocol and 25 participants to an "interviewer-track" protocol. In the respondent-track protocol, participants determined the order of questioning and maintained control of records. In the interviewer-track protocol, interviewers followed a fixed order of questioning and controlled the records during the interview.

Analyses show that participants in the respondent-track group reported significantly more of their expenses using records as compared to the participants in the interviewer-track group. This difference may be explained by two causes: an interviewer effect, whereby one interviewer had significantly less records usage than the other five interviewers; and more use of electronic records in the respondent-track group. And, qualitative feedback from participants emphasized the importance of advance communication for setting participant expectations and understanding of the interview content.

Key Words: interview protocols, records, Consumer Expenditure Survey, interviewers

1. Introduction

The Gemini redesign proposal for the Consumer Expenditure (CE) Survey calls for two separate interviews as well as individual diaries using a single sample of consumer units. After the first interview collects reasonably easy-to-recall expenditure categories and the individual diaries collect small, frequent, and personal expenditure categories, the second interview will collect information about expenditures for categories that we believe respondents are likely to have financial records for, such as mortgages, utilities, and health insurance premiums. For the second interview, respondents are to be encouraged to collect financial records related to these expenditure categories to use when answering the interview questions. Primary records, such as bills and invoices, and secondary records, such as account statements and bank statements showing summaries of expenses, can provide accurate, detailed, expenditure information that is typically difficult for respondents to recall. The goals of this redesigned records interview include maximizing the quality of the data collected while minimizing both the length of time needed for the interview and the burden on the respondent.

Within these goal specifications, there is a range of possible implementations of the records interview. For example, while the Gemini redesign proposal specified that respondents should use records, the proposal did not specify how the interview should be administered, including characteristics such as section order and incorporation of records. Research was needed to evaluate alternative interview protocols' interview length, respondent burden and respondent use of records. Previous studies have explored the feasibility of collecting information from records (Fricker & Edgar, 2010; RTI, 2011), but there has not yet been a test of the field protocols for a records-based interview.

Although quantitative results and statistics are presented here, we make only broad conclusions without interpretation of the magnitude of the effects. We strongly caution against making inferences based on the precise coefficients presented due to the small sample size of the field test. These results are not intended to be representative of the CE sample.

2. Study Design

Six interviewers (three female) with previous experience conducting in-person household interviews were recruited by Westat to administer this study. The average tenure at Westat of the interviewers was three years. The interviewers were trained to use one of the two protocols being tested (three per protocol), and were not made aware of the other protocol. The two interview protocols for testing are described below:

Interviewer-track protocol
In the interviewer-track protocol, interviewers were instructed to go through the spreadsheet-based computer-assisted personal interviewing (CAPI) instrument in a prescribed order, following the current interview protocol. Interviewers were told to review the records brought to the interview by the participant before starting the interview, and to maintain control over the records during the interview, including locating records and expenditure information. If the participant expressed a preference to handle the records, then the participant was allowed to do so.

Respondent-track protocol
In the respondent-track protocol, interviewers were instructed to ask the respondent for input on how to proceed through the interview. The participant directed the order of the interview sections and maintained control of all records. After the participant had completed reporting the expenditures related to the records brought to the interview, the interviewer would then return to any sections with missing items; the interviewer would proceed through the interview asking the participant about any expenses that had not yet been discussed.

3. Results

A total of 52 participants took part in this study, with 27 participants assigned to the respondent-track protocol and 25 participants to the interviewer-track protocol. Expenditure information and participant demographics were collected using a spreadsheet-based CAPI instrument. In addition, the instrument also collected paradata, including timing and some interviewer navigation movements. Interviewer and participant debriefing instruments collected further information regarding checklist usage and burden.

3.1 Participant Demographics
Participants were assigned to protocol groups to approximate similar demographics across the two groups. Household size did not vary by protocol track, with mean household sizes of 2.1 people in the interviewer track group (SD = 1.3) and 2.2 people in the respondent track group (SD = 1.1; $t(50)$ = .44, ns). Reference person education did not vary by protocol group ($X^2(5)$ = 2.7, ns). Additionally, age of reference person did not vary by protocol group ($F(1, 50)$ = 0.14, ns). There was a significant difference in combined yearly income and monthly social security income between protocol groups ($F(1, 43)$ = 6.06, p = 0.018; Mann-Whitney U = 322.00, n_1 = 25, n_2 = 19, p = .045) with participants in the respondent-track reporting higher incomes. Income is included in subsequent models for analysis though the protocol group difference is likely due to chance given the small sample size used in this analysis (n = 44).

3.2 Expenditure reporting and record use
A summary of average expenditure reporting and record use is shown in Table 1 below.

Table 1: Summary of expenditure reporting and record use, by protocol group (n = 52)

Measure	*Interviewer-Track*		*Respondent-Track*	
	Mean	*SD*	*Mean*	*SD*
Mean count of records used	22.32	15.23	27.70	15.11
Mean count of expenditures	55.32	12.25	53.93	13.19
Mean proportion of records used	38.46%	20.65%	49.39%	20.47%

Proportion of expenses reported using records
Analysis of the proportion of expenses reported using any records (primary, secondary, or other) shows that participants in the respondent track reported significantly more of their expenses using records as compared to the participants in the interviewer track (respondent-track protocol: mean = 49.4%, SD = 20.5%, Min = 3.0%, Max = 79.0%; interviewer-track protocol: mean = 38.5%, SD = 20.7%, Min = 7.0%, Max = 88.0%). However, this difference between protocol groups may be explained by an interviewer effect, whereby one interviewer in the interviewer-track protocol had less records usage than the other five interviewers (see Table 2). Once this interviewer is accounted for in a linear regression analysis, protocol group no longer predicts records use. This interviewer effect may be driven by differential interviewer behaviors in encouraging participants to use records or to search for additional records during the interview visit. See Table 3 for these results.

Table 2: Proportions of expenditures reported using records, by interviewer (n = 52)

Protocol Group	Interviewer	Number of Participants	Mean	SD	Min	Max
Interviewer-track	Interviewer 1	11	25.0%	6.8%	14%	35%
	Interviewer 2	5	54.9%	25.2%	26%	88%
	Interviewer 3	9	45.9%	20.5%	7%	71%
Respondent-track	Interviewer 4	9	45.7%	25.9%	3%	71%
	Interviewer 5	8	49.6%	18.2%	17%	70%
	Interviewer 6	10	52.5%	18.3%	18%	79%

Table 3: Linear regression results predicting proportion of expenses reported using records (n = 44)

Predictor	B	SE	t	p
Protocol group	.031	.062	.502	.619
Household size	.012	.025	.482	.633
Estimated annual income (centered)	.000001	.000	1.598	.119
Age of reference person (centered)	.003	.002	1.448	.156
Time spent collecting records before visit	.000	.000	.575	.569
Interviewer effect, dummy (Interviewer 1 = 1)	-.246	.069	-3.592	.001

Total number of expenses reported

The mean number of expenses reported across all participants was 54.6 expenses (SD = 12.7, Min = 30, Max = 82). A regression model including all factors expected to correlate with total number of expenses reported finds that only household size has a significant effect on total number of expenses reported (see Table 4). In other words, larger households reported more expenditures, even when controlling for other factors. Protocol group did not have a significant effect on the total number of expenses reported; section level differences between protocols in total number of expenses reported was not analyzed.

Table 4: Linear regression results predicting total number of expenses reported (n = 44)

Predictor	B	SE	t	p
Protocol group	-3.232	3.844	-.841	.406
Household size	3.426	1.678	2.041	.048
Estimated annual income (centered)	5.684E-5	.000	1.231	.226
Age of reference person (centered)	.107	.115	.930	.359
Time spent collecting records before visit	.005	.019	.273	.786
Proportion of expenses using records	17.628	9.370	1.881	.068

Overall, the analyses examining expenditure reporting and record use suggest that the respondent track protocol may be more successful than the interviewer track in encouraging the use of records to report expenditures. The significant effect of the single interviewer also indicates the importance of interviewer training and understanding of the purpose of records usage. Given that records use lengthens the interview, interviewers may deliberately seek to reduce the number of records used during the interview.

3.3 Record use during the interview

Interviewers reported that participants used an average of 2.5 record types (e.g., paper bills) out of 6 possible types during the interview (SD = .94, Min = 1, Max = 5). The frequency of each source type is shown in Table 5. The source most frequently reported as being used was paper bills (86.5%). Overall, 92.3% of participants used paper records and 53.8% used electronic records. However, participants in the respondent-track protocol group were significantly more likely to use electronic records during the interview (respondent-track protocol: mean = 70.4%; interviewer-track protocol: 36.0%; $t(50) = 2.60$, $p < .01$). This is likely in part due to the logistical difficulties involved in interviewers accessing online accounts through the participants' computer or tablet. Interviewers were not given any special instructions on how to ask for or handle electronic records, other than to remind participants before and at the start of the interview that some records may be electronic; participants in the interviewer-track were also able to access their records themselves rather than handing over control to the interviewer, if they preferred.

Table 5: Record sources used during the interview (mark all that apply) (n = 52)

Record source	Percent of all participants	Percent in Interviewer-Track	Percent in Respondent-Track
Paper bills	86.5%	84.0%	88.9%
Paper bank/credit card statements	51.9%	68.0%	37.0%
Other paper documents	36.5%	48.0%	26.0%
Any paper records	*92.3%*	*92.0%*	*92.6%*
Records accessed with a laptop or computer	46.2%	28.0%	63.0%
Records accessed with a smart phone	15.4%	16.0%	14.8%
Records accessed with a tablet	1.9%	0.0%	3.7%
Any electronic records	*53.8%*	*36.0%*	*70.4%*

3.4 Interview Length and Respondent Burden

Interview length

The mean length of the interview overall, from the start of the household roster to the end of the final survey section, was 62.9 minutes (SD = 16.0, Min = 37.7, Max = 92.6). Analysis of interview length shows that protocol group did not have a significant effect (respondent-track protocol: mean = 64.7, SD = 14.7; interviewer-track protocol: mean = 60.9, SD =

17.3), while the proportion of expenses reported using records has a positive effect (see Table 6). The total number of expenses reported also has a marginal positive effect. In other words, use of more records increases the length of interviews, accounting for the number of expenditures reported regardless of protocol group. Additionally, the use of electronic records does not significantly contribute to the length of the interview.

Table 6: Linear regression results predicting interview length (minutes) (n = 44)

Predictor	B	SE	t	p
Protocol group	.790	5.045	.157	.877
Household size	.649	2.167	.299	.766
Estimated annual income (centered)	.00005	.000	.753	.457
Age of reference person (centered)	.008	.154	.049	.961
Time spent collecting records before visit	.019	.024	.794	.433
Total number of expenses reported	.405	.207	1.959	.058
Proportion of expenses reported using records	31.438	13.383	2.349	.025
Use paper records	9.083	8.254	1.100	.279
Use electronic records	-2.158	5.573	-.387	.701

Respondent burden
During the self-administered post-interview debriefing, respondents were asked to rate how burdensome they felt the interview experience was on a scale of 1 (not at all burdensome) to 4 (very burdensome). The mean overall rating was 1.4 (SD = 0.70, Min = 1, Max = 4). The mean response to the same question but from the perspective of how "other people" would feel was 2.1 (SD = 0.90, Min = 1, Max = 4). These ratings did not differ between protocol groups (median test, ns) and the distributions of responses are not significantly different between the two protocol groups for either question (self: $X^2(3) = 1.4$, ns; other: $X^2(3) = 0.17$, ns). The most frequently provided response to the question about the respondent's own experience was "not at all burdensome" (1), while the most frequently provided response to the question about other people was "somewhat burdensome" (3). Despite these feelings of burden due to the interview, there is no evidence of a difference between protocol groups.

3.5 Records Collection Behaviors
The time spent gathering records before the visit was collected as an open-ended response during debriefing at the end of the study visit and these responses were then converted by the researchers into quantitative measures. The time spent does not vary by protocol, as expected because there were no differences between groups before the interview visit (respondent-track, mean = 90.4 minutes, SD = 100.4, Min = 10, Max = 420; interviewer-track, mean = 100.3, SD = 90.2, Min = 5, Max = 306). Overall, a regression model for time spent gathering records including protocol group, household size, income, and age of reference person did not find any significant factors. These results suggest that time spent in advance of the interview gathering records did not vary on any participant demographic factors. Moreover, these results lower the likelihood that any differences in expenditure reporting and records use were due to records collection behaviors rather than the

protocols. It is also possible participants did not accurately recall the time spent gathering records in response to this debriefing question, and instead gave universally reasonable and socially desirable responses.

Analysis of the effect of time spent gathering records on the total length of the interview finds no significant association ($r^2 = .115$, ns). Likewise, analyses of the effect of time spent gathering records on total number of expenses and proportion of expenses reported using records show no relationships (total expenses: $r^2 = .088$, ns; using records: $r^2 = -.003$, ns). In other words, there is no evidence that time spent gathering records before the visit has any relationship with measures related to participant expenditure reporting or their interview experience.

Based on self-report, 92% of all participants used the checklist during their record gathering activity before the interview visit. This behavior did not vary by protocol group, as expected ($t(48) = 1.21$, ns). When asked how burdensome it was to collect the records in advance, overall participants gave a mean rating for themselves of 1.79 (SD = .89) and a mean rating for "other people" of 2.52 (SD = .93). There were no differences between the protocol groups (median test, ns).

4. Discussion

Overall, this analysis finds that both interview protocols can be used to successfully collect expenditure data. Participants in both protocols used records during the interview to report their expenses. There are no significant differences between the tested protocols in regards to interview length or respondent burden. However, the use of records does significantly lengthen the interview. It was not within scope of this analysis to investigate whether the use of records increases data quality.

Lessons learned: Interviewers
Interviewers reported feeling sensitive about handling respondent records, during training even before data collection had begun. Although the content of the Consumer Expenditure survey already includes potentially sensitive information, records add a new source of private information that the interviewers in this study did not feel comfortable taking from participants. Interviewers who regularly collect data for the Consumer Expenditure Survey may be more acclimated to handling respondent records, but we expect that this sensitivity regarding respondent records may be common to all interviewers. In a survey of interviewers (n = 717), one question asked interviewers how they would prefer to handle respondents' financial records and why. In addition to concerns related to interview length and data accuracy, responses included concerns about identity theft, invasion of respondent privacy, and personal unease (Kopp et al, 2016).

We also learned from this study that respondent use of records depends in part on encouragement from the interviewers. Because record use adds length to the interview, which is a tangible cost to the interviewer, interviewers must be incentivized to encourage use of records during the interview.

Lessons learned: Respondents
This study showed that record sharing presents logistical difficulties if respondents access electronic records on demand during the interview; the respondents cannot be expected to share their account information with the interviewer.

Participants in this study also showed a preference for letting the interviewer control the administration of the interview rather than taking control themselves.

Lessons learned: Survey content
The checklist is a key tool for introducing the content of the records interview and setting respondent expectations. As such, the checklist should include all the guidance necessary for respondents to gather records (e.g., including explicit information about collecting income records for the past 12 months) and organize them according to section (rather than chronologically by date or by any other system). Further testing is needed to understand whether the checklist content should identify the expenditures to be collected during the interview or the records on which those expenditures are likely to be found. Preliminary feedback indicates that naming only the records may result in miscommunication and confusion over what information is required. Interviewers reported that participants were sometimes not well-prepared for the content of the interview, resulting in frustration with the advance communication and time spent during the interview visit to search for and gather additional records.

A hybrid approach in which respondents maintain control of their records, a protocol associated with increased electronic records usage and reduced respondent unease, while interviewers control the interview and the order of sections is recommended for further testing.

Acknowledgements

References

Fricker, S. & Edgar, J. (2010). Exploring sources of error in the Consumer Expenditure Survey: Results from a small-scale validation study. Paper presented at the 65th annual conference of the American Association for Public Opinion Research, Chicago.

Kopp, B., Safir, A., Jilk, L, Tan, L, & Abdirizak, S. (2016). 2015 Field Staff Survey Report. Internal BLS report.

RTI (2011). U.S. Consumer Expenditure Records Study: Final report.