Examining Changes in Filter Question (FQ) Reporting in the Consumer Expenditure Quarterly Interview Survey October 2013

Brett McBride[1]

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

Abstract

The survey design involving a screener or filter question followed by a series of more detailed questions is used in many surveys. A concern in these surveys is that respondents may learn that reporting a certain answer to a filter question will extend the interview through a series of follow-up questions and thus will alter their responses in a way that avoids their exposure to follow-up questions. Alternatively, the cumulative cognitive burden experienced by respondents after answering many survey questions may be a principal factor in respondent satisficing, and thus measurement error, toward the end of surveys. The exploratory research described in this paper identifies changes in response patterns to filter questions in the Consumer Expenditure Quarterly Interview Survey (CEQ). A model is used to examine the role of respondent characteristics and interview burden measures on declining filter question endorsement rates. Aside from interview content, this research found that respondent characteristics impacting cognitive burden had effects on a declining trend in filter question endorsement rates.

Key Words: Respondent conditioning, Consumer Expenditure Survey, paradata, data quality, questionnaire design

1. Introduction

Survey research has found that the process of responding to questions may lead to respondent conditioning, or changes in how individuals' respond. A significant amount of research has been carried out on panel conditioning, the impact of multiple interviews on respondent behavior, typically finding little effect of completing multiple interviews on responses. Research conducted on the Bureau of Labor Statistics' Quarterly Interview (CEQ) Survey has similarly found little evidence of panel conditioning on expenditure reports (Shields & To, 2005; Yan & Copeland, 2010). A main focus of this article is to investigate whether respondent conditioning affects responses within a single CEQ interview.

The CEQ is the component of the Consumer Expenditure (CE) Survey that collects retrospective expenditure information through in-person or telephone interviews. Questions on the CEQ ask many details about a household's expenditures using a questionnaire structure in which an initial Filter Question (FQ) determining if an expenditure occurred is followed by a series of more detailed questions, a pattern known as an 'interleafed format.' This structure of a FQ with detailed follow-up questions is repeated in a cycle for the range of expenditure categories asked about throughout the interview. This interleafed format may lead to measurement error if respondents 'learn' that responding 'yes' to a FQ will result in a series of follow-up questions, and thereafter decide to answer FQs in a way that limits the follow-up questions that they have to answer. This phenomenon, referred to as 'motivated underreporting,' may occur within as well as across interviews.

Research has found evidence that the interleafed format has led to conditioning. One experimental study administered two interleafed format sections, but rotated the section order. Their finding, of borderline significance, was that the section coming first always had higher FQ endorsement, suggesting measurement error due to the question format (Jensen et al., 1999). Another method used to measure whether respondents learn of the FQ-follow-up cycle has been to design a 'grouped' format that administers all FQs at the beginning and then asks all follow-up questions based on the earlier FQ responses. This format is contrasted with the traditional interleafed FQ format. Employing this method in a split-ballot telephone interview, Kessler and colleagues (1998) found lower FQ endorsement rates in the interleafed format than in the grouped format, a finding replicated using in-person interviews (Duan et al., 2007). Similar studies have examined how the interleafed format affects responses to CE Survey questions, resulting in mixed results (Bosley et al., 1999, Kreuter et al., 2011). Kreuter and colleagues found higher endorsement rates for the grouped format, but cautioned that higher endorsement was not synonymous with better data quality in some of the sections having validation data. They also noted that respondents gave more "don't know" responses to the follow-up questions in the grouped format, suggesting some trade-offs to focusing on the accuracy of FQ responses.

The survey format is just one factor that may account for changes in response patterns. The mechanism underpinning many causes of respondent conditioning is respondent satisficing. Jon Krosnick and Duane Alwin proposed that respondents practice 'satisficing' to reduce the psychological costs involved when answering survey questions, instead of optimizing during the response process (Krosnick & Alwin, 1987; Krosnick, 1991). Satisficing can range from a slight reduction in the effort needed to supply responses to the respondent making no search of their memory in order to retrieve a response. Satisficing behavior is influenced by the difficulty of the survey task, respondent motivation, and respondent ability. In surveys such as the CEQ, respondents may resort to satisficing due to the cognitive burden imposed by having to answer many questions, and this may be more pronounced among those not motivated to do the survey in the first place, or those with limited cognitive ability. Therefore, changes in response patterns may occur because respondents learn of the interleafed FQ structure and eventually respond in a way that reduces the difficulty of the survey task, or they become fatigued over the course of the survey and lose the motivation to provide accurate responses, or a combination of the two. To definitively attribute changes in response patterns to the interleafed format would require an experimental design. Instead, this research makes use of an observational design that suggests whether the format may be leading to respondent conditioning. Beyond investigating changes in FQ endorsement in the CEQ, this article explores the role of respondents' pre-interview concerns about participating, expected correlates of interview burden (e.g., interview length, amount of expenditures), and other survey characteristics on changes in FQ endorsement.

2. Data and Analysis Methods

2.1 Consumer Expenditure Survey
The U.S. Consumer Expenditure Survey is a national, household survey collecting information on the buying habits of consumers. The Census Bureau collects the survey data on behalf of the Bureau of Labor Statistics. Data from the CEQ as well as from a separate

diary survey are both used to determine expenditure weights in the Consumer Price Index, which measures changes in prices paid for goods and services, and are also used by economists, academics and market researchers. This research involved data from the CEQ, which had a response rate of 71 percent in 2011 (AAPOR, 2011, response rate 1). Data were taken from May 2011 interviews conducted as part of the first of five interviews of participating households. I identified questions on the interview that all respondents would answer (excluding questions asking if a respondent had any other purchases in the same expenditure category) and that determined whether a follow-up expenditure amount would be collected. This resulted in 303 FQs for which respondents' answers would determine which follow-up questions would be administered. As the objectives of the project were to determine changes in response patterns within an interview and how this was affected by the length of the interview, the project removed the 9% of cases in which respondents completed interviews in multiple sessions, resulting in an analysis dataset consisting of 545 interviews conducted in a single session.

The 303 FQs were distributed throughout all but the first of the nineteen sections in the CEQ. The following FQs from section 12 were typical of those asked in the interview – "*Since [month], have you had any expenses for… Driver's licenses? Vehicle inspection? State vehicle registration? Local vehicle registration?*" This question was not analyzed at the question-level (e.g., recording it as being endorsed if any one of the four items resulted in a 'yes' response). Instead, this and similar questions were treated as having multiple, separate FQs, since each item led to detailed follow-up questions (e.g., "In what month did you have this expense," "What was the total amount of the expense."). Respondents who learned of the CEQ's interleaved format and altered their responses were expected to do this in response to each item, making the item-level the appropriate level of analysis for the objectives of this research.

2.2 Analysis Methods

2.2.1 Measuring and Predicting Change in FQ Endorsement

In order to examine changes in FQ endorsement, it was necessary to define a measure of change. I divided the 303 FQs into three groups or 'bins' of approximately 100 FQs each. This division was carried out with the expectation that respondents would become exposed to the FQ-follow up structure in the first bin (corresponding to sections 2 through 6 of the interview[1]), leading to the greatest potential for changes in FQ response patterns by the third bin (sections 13 through 19). It was also expected that if respondents became fatigued from the interview length, this would develop by the last third of the interview. Therefore, the measure of change examined by this project was the average endorsement rate in the third bin minus the average rate in the first bin, henceforth referred to as the FQ trend. A two-way repeated measures analysis of variance (ANOVA) was used to measure the extent of change in endorsement rates throughout the interview, as well as the effect of respondent motivation. A regression model was used to more effectively quantify the impact of respondent motivation on the FQ trend, as well as to control for factors associated with declines in the FQ trend. The linear regression model regressed the change in FQ endorsement from the first to last third of the interview on expected correlates of the FQ trend.

---

[1] No FQs with collected data were in section 1.

2.2.2 Descriptive and Subgroup Analyses to Control for Interview Content

An awareness of the fluctuations in endorsement rates was achieved by dividing the interview into smaller segments for which average endorsement rates were calculated. Endorsement rates in different sections of the interview were examined to identify if there were apparent correlations between interview content and endorsement rates that might explain the FQ trend. To seek to control for the content of FQs being associated with endorsement rates, a subgroup analysis was performed for respondents having similar response patterns.

3. Results

3.1 Filter Question Response Patterns

The 303 FQs examined in this research had an average endorsement rate of 5.0%. This rate was low in part due to the short, one-month reference period, and in part due to the many FQs in the survey that ask about infrequently-purchased items. Average endorsement rates for sequential segments, each of approximately 25 FQs, were calculated and charted (Figure 1). Fitting a polynomial trend line to the chart (in black) revealed a decline in 'yes' responses to FQs toward the middle of the interview, but a later increase.
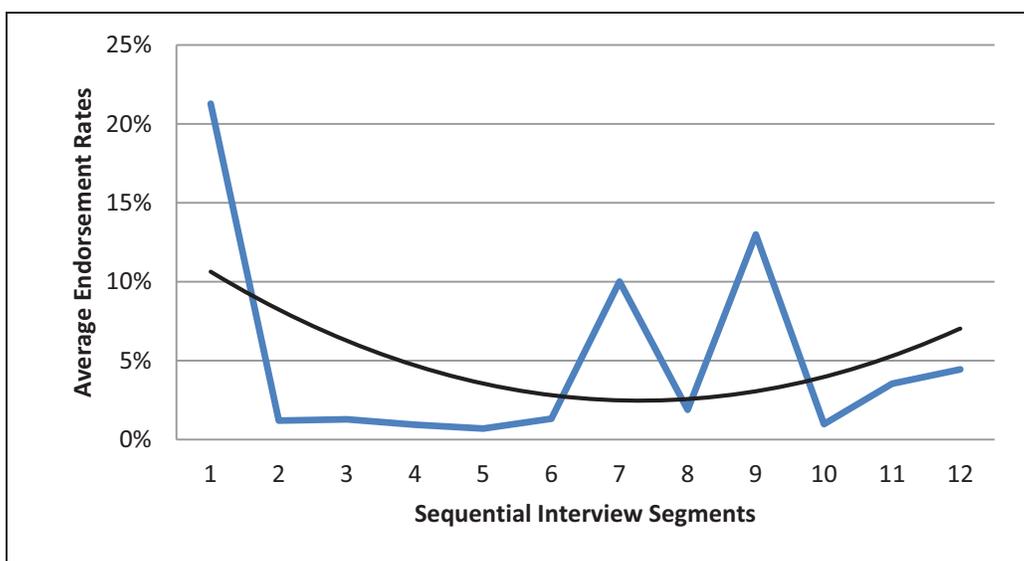
Figure 1: FQ endorsement rates throughout interview (average endorsement rates for 25-FQ segments).

The chart shows a noticeable trough in the early-middle portion of the interview, which corresponds to low endorsement rates in Sections 6 through 9. Section 6 (in bin 1), containing questions about appliance expenditures, had 52 FQs of which only one question had more than 5% of respondents answering 'yes.' In Sections 7 through 9 (in bin 2), which included questions about home maintenance, furniture, and apparel, only 10 of the 73 FQs had endorsement rates exceeding 5%.

3.2 ANOVA

A repeated measures ANOVA was used to determine if there were significant changes in FQ endorsement rate over time (within the interview), and whether changes were correlated with respondent motivation. Time was represented by the three bins for which average FQ endorsement rates were calculated. The main measure of respondent motivation to complete the survey was obtained from paradata collected from Census' Contact History Instrument (CHI). Interviewers collected doorstep concern information from households during the recruitment phase. A dichotomous variable indicating whether a doorstep concern was mentioned prior to the first interview was included in the ANOVA model as a proxy for motivation. Table 1 presents degrees of freedom and F-test statistics from the significance tests.

Table 1: Examination of Time and Doorstep Concerns Effects
(n=540)

Measure	df	F-Statistic
Time Effect	2	204.73***
Concerns Effect	1	0.22
Time by Concerns Interaction	2	0.69
Contrast: Bin 1 & Bin 2	1	394.40***
Contrast: Bin 1 & Bin 3	1	25.1***
*** p<0.001		

The ANOVA test revealed a significant time effect across the three bins of the interview. Contrasts revealed that the difference in endorsement rates between bin 1 and bin 2 (from 6.2% to 3.4%) was significant, as was the smaller change between bin 1 and bin 3 (from 6.2% to 5.5%). This analysis also sought to identify whether the presence of doorstep concerns affected differences in FQ trends. The test for between-subjects effects did not indicate a significant difference between those expressing concerns and those not expressing concerns on average FQ endorsement ($F(1, 538) = 0.22$, $p = .64$), nor was an interaction found between respondents' expression of concerns and their FQ response patterns ($F(2, 1076) = 0.69$, $p = .50$). This analysis revealed that there was a significant decline in FQ endorsement within the interview, although it did not differ based on whether respondents had expressed a doorstep concern.

3.3 Linear Regression Model

The regression model included data from a total of 539 households[2]. The dependent variable in the model was the FQ trend, with negative values indicating a declining trend in response patterns. The dependent variable was not normally distributed, but departures from normality were judged not significant enough to distort the model findings. The change in average FQ endorsement from 6.2% in bin 1 to 5.5% in bin 3 represented an 11% decline. However, splitting the FQs between those in the first and last halves of the interview resulted in an increase in average FQ endorsement from 4.5% to 5.5%.

---

[2] 6 cases were removed through listwise deletion due to missing values.

Several variables were available to control for the demographic differences of respondents completing the CEQ. Respondent age was used as a continuous variable. Education represented the number of years of schooling a respondent reported. This variable was seen as a proxy for respondent cognitive ability in predicting the likelihood of satisficing. A dummy variable was used to indicate respondents living in an owned home, and a variable representing the number of household members was included in the model. The model included a continuous variable representing the percent of a household's total expenditures devoted to necessities[3]. This was considered a proxy measure for poorer households, although it could also encompass households choosing to spend less of their budget on non-essential items.

Five research variables were included in the model. A non-zero value for doorstep concerns indicated the presence of at least one doorstep concern[4]. A measure of burden, labeled 'expenditures,' was a count of the number of non-zero expenditures a household reported. Another measure of burden was the count of "don't know" responses given by the respondent. Since respondents are not given the option of providing a "don't know" response to most FQs, this variable mainly reflected responses to follow-up questions[5]. A non-zero value for the dichotomous variable 'interview mode' indicated that the interview took place in-person. In-person interviews were expected to be less likely to have a declining FQ trend than telephone interviews, since their features – non-verbal cues, a slower pace, and reduced likelihood of respondent multi-tasking – were seen as discouraging satisficing. The model included a time measure of burden, representing the duration of the interview, in minutes, from the start of the interview to the beginning of section 13 (the beginning of bin 3). Table 2 presents the measures of central tendency for these variables.

Table 2: Regression Variable Measures of Central Tendency (n=545)

Variable Description	Mean	Median
FQ Trend	-0.7	-1.0
Education	13.4	13.0
Age	49.5	48.0
Household Size	2.5	2
Household Tenure	0.633	1.000
Prop. Spent for Necessities	40.2	36.6
Doorstep Concern(s)	0.462	0.000
Number of Expenditures	16.9	16.0
Number of DK Responses	0.760	0.000
Bin1 to Bin2 Duration	28.2	26.0
In-Person Interview	0.844	1.000

---

[3] Basic necessities included expenditures for mortgage (or rent), assessments, fees, ground rent, telephones, utilities, Internet, and groceries. This variable was multiplied by 100 to be in the same scale as the dependent variable.

[4] Five respondents with missing values were imputed to have a zero value for the concern variable based on associated variables, in order to be retained in the analysis model.

[5] Only two responses to FQs in the data had a "don't know" response.

The regression model predicting the FQ trend compared a full model against a reduced model. Both models explained 29% of the variance in the FQ trend. The model coefficients and standard errors are presented in Table 3.

Table 3: Regression Model Coefficients Predicting FQ Trend (Bin 1 to Bin 3)

	Full model (n=539)		Reduced model (n=539)	
	Coeff.	SE	Coeff.	SE
Intercept	-5.089***	(0.941)	-5.145***	(0.870)
- Education	0.165***	(0.042)	0.165***	(0.042)
- Age	0.029***	(0.008)	0.028***	(0.008)
- Household Size	-0.157	(0.087)	-0.167	(0.086)
- Household Tenure	-1.719***	(0.295)	-1.745***	(0.292)
- Prop. Spent for Necessities	-0.012*	(0.005)	-0.012*	(0.005)
- Doorstep Concern(s)	-0.245	(0.244)	-0.263	(0.241)
- Number of Expenditures	0.177***	(0.019)	0.170***	(0.017)
- Number of DK Responses	0.146*	(0.075)	0.140	(0.074)
- Interview Duration	-0.009	(0.010)		
- Interview Mode	0.003	(0.340)		
R^2		0.292		0.290

*p<.05; **p<.01; ***p<0.001

In the full model, four of the variables did not attain significance (though household size was borderline significant in both models (p<0.10)). The mode of the interview was found to be highly insignificant when controlling for other model variables (p=0.99), indicating the FQ trend was unaffected by whether or not the first interview was conducted in-person. For every additional minute of elapsed time up until the start of bin 3, there was an associated 0.009 percentage point decrease in average endorsement rates, though this coefficient was not significant (p=0.35). Although also not significant, the variable indicating whether a respondent expressed one or more doorstep concerns was retained in the final model, as it was a main analysis variable. This variable in the full model indicated that respondents expressing an initial concern were associated with a 0.25 percentage point decrease in FQ trend (p=0.32). The reduced model coefficients were largely unchanged from those of the full model, though the coefficient on the variable for the number of "don't know" responses became of borderline significance (p=0.059). Model diagnostics indicated that the model assumptions were largely met.

3.4 Controlling for Interview Content

3.4.1 Descriptive Analysis

Although the ANOVA test indicated a declining FQ trend, I sought to identify the extent to which this was an artifact of differences in the frequencies of expenditure category reports in the three bins that were examined. To do this, the project examined annual 2011

CE estimates indicating the quarterly percent of households that reported expenditures in various CEQ categories[6]. Reporting frequencies were not available for all sections (e.g., the proportion reporting ownership of a vehicle, Section 11) nor were they aligned with CEQ sections for others (e.g., 'reading' is only a subset of the expenditures respondents report in Section 17). An external truth measure, had it been available, would have provided a more appropriate comparison, as its data would be unaffected by the same response error patterns that were the subject of investigation in this research project. Those limitations notwithstanding, this method capitalized on the data that were available in order to provide a general picture of whether the expenditure categories of the three bins provided comparable reporting frequencies. Table 4 provides FQ characteristics and publication data for each of the examined interview sections.

Table 4: Comparison of Bin Characteristics: FQ Typologies and 2011 Reporting Frequencies from Publication Tables

Bin	Section	No. FQs	Month 6 Mean FQ endorsement	Publication Table Category	Qtrly % Endorsing
	2	1	63%	Owned dwellings	65%
	3	5	3%	Other lodging	19%
1	4	18	25%	Telephone services	92%
	5	24	1%	Maintenance and repair services	1%
	6	52	1%	Major appliances	9%
	7	11	1%	[no information in publication table]	
	8	36	1%	Furniture	12%
2	9	26	6%	Apparel and services	75%
	10	1	4%	Leased & rented vehicles	6%
	11	1	85%	[no information in publication table]	
	12	27	3%	Vehicle maintenance & repairs	55%
	13	6	30%	Homeowners insurance	23%
	14	1	67%	Health insurance	64%
	15	34	2%	Medical services	44%
3	16	9	3%	Education	15%
	17	13	3%	Reading	37%
	18	1	2%	[no information in publication table]	
	19	37	5%	Miscellaneous	41%

It can be seen that there was a similar range in the quarterly reporting frequencies (shown in the last column) for bins 1 and 3 – from 1% to 65% in bin 1 and from 15% to 64% in bin 3. Examining the quarterly reporting frequencies, significant proportions of respondents reported having expenditures in categories such as insurance and medical services which are asked about in the last third of the interview. As there was no clear difference in reporting frequencies between bins, the FQs in bins 1 and 3 were seen as sufficiently comparable for the purpose of measuring changes in response patterns. The

---

[6] CE data is designed to provide quarterly estimates, and so monthly equivalents of these reporting frequencies are not available (only the first interview involves use of a monthly reference period).

third column indicated, at a section-level, the number of FQs and average FQ endorsement rates. As expected, these section-average FQ endorsement rates show some relation to the published reporting frequencies. However, monthly endorsement rates did tend to be lower than quarterly reporting frequencies. There appeared to be a correlation between the level of detail at which FQs are asked in certain sections and large gaps between FQ endorsement rates and reporting frequencies (e.g., 4, 8, 9, 12, 15 and 17). This suggests that, although respondents may make expenditures within these categories, the numerous detailed items about which the expenditures are asked may have a depressing effect on FQ endorsement rates.

3.4.1 Subgroup Analysis

Attributing changes in response patterns across the interview to respondent satisficing was complicated by the possibility that, in later sections, respondents did not actually have expenditures to report, explaining observed declines. As a result, a separate control analysis was carried out that sought to control for the confound of the location of expenditure categories on the FQ trend. This intended to identify sections where it would be unlikely for certain respondents not to have expenditures, based on CE publication data. Respondents not reporting expenditures in these sections were examined to see if they differed from comparable respondents who did in ways consistent with the research hypothesis (e.g., more burden and reluctance). This subgroup analysis was performed on the subset of 109 respondents who were over the age of 65. CE Publication tables indicated that over 95% of households with reference persons in this age category reported having health care expenses (as collected in Sections 14 and 15) per quarter in 2011. The differences in trend values and related measures by these respondent groups are presented in Table 5.

Table 5: FQ Trend and Correlates by Health FQ Endorsement
(Among Respondents Older than 65, n=109)

	Health FQ endorsed (n=87)	No health FQ endorsed (n=22)	Difference	T value
FQ Trend (bin 3 - bin 1)	0.4	-3.4	3.8	5.5***
FQ Trend (without health FQs)	-1.7	-3.4	1.7	2.7**
Proportion expressing concern	45%	32%	13%	1.1
Duration (bin 1 through bin 2)	30.8	20.1	10.7	3.3**

**p<.01; ***p<0.001

Respondents who endorsed one of the 35 health FQs in Sections 14 and 15 had a positive overall FQ trend, compared to those who did not report any health expenditures (0.4 percentage points and -3.4 percentage points, respectively). Since some of the negative trend value among respondents in the latter group was attributable to their not reporting health expenditures, the trend was recalculated, removing health FQs from analysis. If all of the difference between these groups was attributable to their health FQ responses alone, no difference would remain after excluding the health FQs from the analysis. The resulting trend indicated that older respondents not reporting any health expenditures still had a significantly larger decline in FQ endorsement than other older respondents (p=0.007). These findings mirrored those of the prior analyses in showing a declining trend in FQ endorsement, and provided further evidence that satisficing may have caused this decline.

Examining the expected correlates of satisficing (e.g., burden – in survey time, and motivation – expressed in doorstep concerns), the expected patterns were not revealed, however. Respondents who did not report health expenditures were less likely to have reported a doorstep concern (though this was not significant), and their interviews were significantly shorter through Section 13. Thus there was little evidence found to suggest low-motivation or burden factored into their reduced reporting.

4. Discussion and Conclusions

These analyses provided evidence of decreased reporting of expenditure categories at FQs over the course of the CEQ interview, when dividing the interview into three segments. However, a more detailed view of the data revealed a significant, non-linear trend in endorsement rates, with the large decline after the first third of the interview largely offset by a rebound in endorsement rates by the last third of the interview. An alternative division of the interview into two segments would not have indicated any decline in endorsement rates. Therefore, these patterns caution against concluding that sections coming later in the interview will automatically have lower FQ endorsement rates, with FQ trends likely to be affected by both question content and placement.

Some measures of respondent characteristics may have been proxies for interview content. The location of FQs dealing with home ownership in the first third of the interview could explain the counterintuitive finding that respondents who owned their homes had a larger decrease in FQ trend than other respondents. Similarly the negative association for the proportion spent on necessities and FQ trend likely represented more items asking about necessities in the first third of the interview. The positive association between age and FQ trend may have resulted from questions dealing with health expenditures in the last third of the interview. Removing the effect of those questions revealed that respondents over the age of 65 had a declining trend. The positive association between years of education and FQ trend was consistent with findings by Krosnick and colleagues that respondents who have greater cognitive ability will be less likely to satisfice when responding to questions.

There was not a significant association between the declining FQ trend and pre-interview intentions toward participating as indicated in both the ANOVA test and regression model. The regression model analysis did suggest that respondents expressing a doorstep concern were associated with a quarter percentage point decrease in FQ endorsement rates from the first to third interview segments, although this finding was not significant. There was an association between the FQ trend and what were expected to be correlates of interview burden – interview length and the number of reported expenditures – but the association was not as anticipated. In the case of time, the association was not significant. In the case of expenditures and the number of "don't know" responses, the significant association was positive. In terms of time, respondents who devoted more time to the survey may be answering more questions and reporting more expenditures, and thus the time spent on the survey may reflect motivation. This would be consistent with the findings from the subgroup analysis of shorter interview durations for older respondents who had not reported any health expenditures. However, the negative sign for the time variable in the regression model did suggest that longer interviews, controlling for other factors such as the number of expenditures reported, may lead to reductions in FQ endorsement rates. It is unclear what role interviewers played in compensating for indications of respondent fatigue (e.g., by speeding up the interview), and thus minimizing any association between

interview duration and FQ trend. Without further research, it is also unknown what role interviewers may play in the non-significant finding of interview mode and FQ trend. In terms of "don't know" responses, the unexpected finding that respondents providing more "don't know" responses were associated with an increasing FQ trend may reflect greater effort among those respondents. Respondents who provided a "don't know" response in these interviews may have been making an effort to report the presence of expenditures at FQs despite not being able to fully report associated expenditure details in follow-up questions. This is consistent with the finding by Kreuter and colleagues that FQs in a grouped format had more "don't know" responses, suggesting that these responses are indicative of more valid overall data. Household size was one measure pointing to cognitive burden's negative impact on the FQ trend. In a finding of borderline significance, for each additional member in a respondent's household, there was an associated 0.16 percentage point decline in the FQ trend. Thus the need for respondents to report by proxy for other household members may be a more significant measure of burden than interview duration or the number of expenditures reported.

4.1 Limitations

The main limitation of this research was the inability to disentangle the FQ trend from section content, as could be achieved through an experimental design rotating the administration of interview sections. Such experimental studies are better equipped to investigate whether respondent conditioning is resulting from the interleaved format, a possibility that this research could not rule out. To mitigate the impact of the observational design of the research, steps were taken to identify what role section content might have played in the FQ trend that was observed. This project accounted for the potential of an actual absence of respondent expenditures in later sections in two ways. First, publication data were used to show that reporting rates for expenditure categories later in the interview were not noticeably lower than for those earlier in the interview, leading to no expectation of a downward bias in FQ trend. Second, sections later in the interview for which respondents older than 65 were highly likely to actually have expenditures were examined. Respondents lacking expenditure reports in these sections were associated with a significantly larger overall reduction in FQ trend, as would be expected if respondent conditioning were present. This difference remained once the effect of FQ responses in that section was removed (i.e., by excluding those questions). Another caveat to the research findings was that project data only covered the one-month reporting period found in the first interview, and thus a smaller sample, with less variability in responses. Focusing on data from the first interview, however, permitted analysis of response pattern trends to be unaffected by the dual effects of respondent attrition and increased respondent familiarity with the instrument.

Other limitations involved the strength of assumptions about how the interview is conducted. As an example, it was assumed that respondents are asked to provide 'yes' or 'no' responses separately for each FQ. However, for many FQs, such as the vehicle question described earlier, interviewers may read all items in a group and then ask the respondent to indicate which, if any, they had purchased. Alternatively, interviewers may tell respondents to refer to supplemental printed information listing the expenditure items but not read the expenditure items out loud. The presence of these 'silent filter questions' may influence how much effort respondents exert in responding to FQs. As the CEQ instrument allows sections to be administered in non-sequential order, there is the possibility that data from later sections may be collected at the early stages of the interview. Although this would be an additional confound to analyzing the impact of cumulative

interview burden on response patterns, anecdotal evidence suggests that interview sections are usually administered in sequential order.

4.2 Conclusions

As the Consumer Expenditure Questionnaire had a declining trend in the endorsement of filter questions, albeit not at a uniform rate throughout the interview, there is the potential that respondents are learning of the interleaved format and that this is affecting their responses. This research suggested that the cognitive burden faced by certain respondents – those with lower education or needing to report on behalf of larger households – was a factor affecting the declining FQ trend. Surveys such as the CEQ may benefit from interviewers using motivational messages mid-way through the interview for respondents who are vulnerable to reduced reporting – those taking longer to answer questions, and those needing to proxy report. The research provided less evidence that the measures of burden associated with long interviews or that respondent motivation, as measured by pre-interview concerns about participation, led to a decreased FQ trend. Due to the observational design of the research, there was inconclusive evidence that within-interview respondent conditioning would be limited by revising the CEQ's interleaved format to a grouped format. As designed, this research suggested that interview content (e.g., the location of housing FQs or FQs dealing with necessities), as much as respondent characteristics, had an impact on declining FQ response trends. This awareness highlights the importance of carefully planning section ordering when designing questionnaires; specifically the need to avoid grouping long stretches of questions that ask about expenditures that respondents did not make, as this may lead to respondent frustration. Similarly, survey designers should be cautious about placing hard-to-recall questions at the end of an interview, as this is a point where respondent cognitive effort may be waning. Further research that employs a more complex method, different burden measures and larger sample sizes may provide a fuller picture of the presence of respondent conditioning on lengthy, interleafed-format surveys.

References

American Association for Public Opinion Research (AAPOR). 2011. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th ed. AAPOR.

Bosley, John, Monica Dashen, and Jean Fox. 1999. When Should We Ask Follow-up Questions about Items in Lists? JSM Proceedings. Survey Research Methods Section. Alexandria, VA: American Statistical Association, pp. 749-754.

Duan, Naihua, Alegria, Margarita, Canino, Glorisa, McGuire, Thomas, and David Takeuchi. "Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats." *Health Services Research* 42, no. 2 (April 2007), pp. 890-907.

Jensen, Peter S., Watanabe, Henry K., and John E. Richters. "Who's up first? Testing for Order Effects in Structured Interviews Using a Counterbalanced Experimental Design." *Journal of Abnormal Child Psychology* 27, no. 6 (December 1999), pp. 439-445.

Kessler, Ronald C., Wittchen, Hans-Ulrich, Abelson, Jamie M., Mcgonagle, Katherine, Schwarz, Norbert, Kendler, Kenneth S., Knäuper, Bärbel, and Shanyang Zhao. "Methodological Studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS)." *International Journal of Methods in Psychiatric Research* 7, no. 1 (February 1998), pp. 33-55.

Kreuter, Frauke, McCulloch, Susan, Presser, Stanley, and Tourangeau, Roger. "The Effects of Asking Filter Questions in Interleafed Versus Grouped Format." Sociological Methods and Research 40, no. 1 (February 2011), pp. 88-104.

Krosnick, Jon A. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5, no. 3 (May/June 1991), pp. 213–236.

Krosnick, Jon A., and Duane F. Alwin. An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly* 51, no. 2 (Summer 1987), 201-219.

Shields, Jennifer, and Nhien To. 2005. Learning To Say No: Conditioned Underreporting in an Expenditure Survey. JSM Proceedings. Survey Research Methods Section. Alexandria, VA: American Statistical Association, pp. 3963-3968.

Yan, Ting, and Kennon Copeland. 2010. Panel Conditioning in the Consumer Expenditure Quarterly Interview Survey. JSM Proceedings. Survey Research Methods Section. Alexandria, VA: American Statistical Association, pp. 3618-3628.