Response Burden: What Predicts It and Who is Burdened Out?

Scott Fricker[1], Ting Yan[2], Shirley Tsai[1]
[1]Bureau of Labor Statistics, 2 Massachusetts Ave., NE Washington, DC 20212
[2]Institute for Social Research, University of Michigan, 426 Thomspon Street, Ann Arbor, MI, 48106

Abstract
Concerns about the burden that surveys place on respondents have a long history in the survey field. As early as the 1920s, survey researchers and practitioners expressed concern about the potential negative impacts of response burden. However, a review of the response burden literature reveals that conceptualizations and measures of burden are still underdeveloped, and as a result findings from empirical research in this area remain equivocal. The Consumer Expenditure Quarterly Interview Survey includes survey questions to measure respondents' perceptions of burden and their attitudes and perceptions about the survey. We examined these data using structural equation modeling in an attempt to address some of the gaps in this literature, focusing especially on the survey characteristics and respondent attitudes affecting perceptions of burden. We found that respondents' subjective perceptions of the survey task had significant direct impacts on burden. Survey features that commonly have been used as standalone proxy measures of burden (e.g., length of interview) had no direct impact, but contributed significantly to the model through their joint effects on subjective assessments of the survey and burden. Finally, we found a significant negative association between respondent motivation and burden. These findings support the notion of burden as a subjective, multidimensional phenomenon, and underscore that the impact of any given survey feature on burden will vary across respondents on the basis of measurable subjective reactions and attitudes.

Key Words: Response burden, Consumer Expenditure Survey, structural equation modeling

1. Introduction

1.1 Background

Survey researchers have long speculated that the burden respondents experience in a survey may negatively impact response rates and data quality (Chapin, 1920; Bradburn, 1978; Sharp and Frankel, 1983). In recent years, falling response rates and the establishment of regulations to reduce the time and effort required of government survey respondents have contributed to a growing empirical literature on this topic. A review of this literature reveals some suggestive findings indicating the negative impact burden can have on survey outcomes, but the results are by no means unequivocal (see Fricker, Gonzalez, and Tan, 2011 for a review).

Given the seemingly plausible and enduring hypothesized effects of response burden, how might we account for these mixed findings? In his seminal article on burden, Bradburn (1978) points to the lack of conceptual development and careful empirical study on this topic as major impediments:

> "[I]t is because of their [survey researchers'] day-to-day concern for the potential burden that they place on respondents that there is little self-conscious research on the issue. It is so much a part of everyday practice, that it is not seen as a topic in need of research" (Bradburn 1978, p35).

Nearly four decades later, our understanding of the causes and consequences of burden remains inadequate. Researchers and practitioners too often rely on loose definitions of *burden,* or continue to employ interview length as a proxy measure of burden despite demonstrable weaknesses with this approach. Bradburn's (1978) concept of burden was multidimensional and reflected the influences of interview length, effort required of respondents, the frequency of interviews, and the amount of stress on respondents. He emphasized that burden is a *subjective phenomenon*, "the product of an interaction between the nature of the task and the way it is perceived by the respondent" (pg. 36). Bradburn suggested several possible factors that could influence respondents' perceptions of the survey task (e.g., interest in survey topic), but only a handful of studies have tried to assess respondents' attitudes and subjective reactions and then examine their impact on burden (e.g., Sharp and Frankel, 1983; Hedlin et al., 2005; Fricker et al., 2011).

To our knowledge, only one study (Fricker, Kreisler, and Tan, 2012) has measured respondents' subjective survey experiences, modeled those data to explore the latent factors thought to underlie perceptions of burden, and then examined associations between those factors and respondents' self-report burden. The authors found support for Bradburn's (1978) conceptualization – respondents' perceptions of survey length, effort and frequency all contributed significantly to perceived burden. The results of this study demonstrate the utility of asking even a small number of burden-related questions at the end of a survey, and illustrate the insights that can be gained from applying multivariate statistical techniques to such data in order to understand how different factors can differentially impact burden dimensions.

1.2 Study Objectives

The aim of the present paper is in part to extend this approach to an analysis of burden in a larger, federal household survey – the U.S. Consumer Expenditure Survey (CE). We also explore a conceptualization of burden that explicitly models the direct and indirect effects of survey features, respondent characteristics, and respondents' perceptions of the survey on burden, in hopes of shedding light on which factors (or combination of factors) are most likely to result in respondent burden.

One of the challenges researchers studying this topic face is the lack of available data in most production surveys that are related to the cognitive, attitudinal, and motivational elements underlying burden. Ideally, one wants information about a vector of respondent characteristics and survey features, to explore interactions between these factors, and their impact on burden (Haraldsen, 2004).

In this study, we take advantage of post-survey questions in the CE that assess respondents' perceptions about their survey experience, and examine other respondent characteristics related to task difficulty and motivation (e.g., household size, their attitudes towards the survey). Our conceptual model also incorporates data available in CE on interviewers' perceptions of respondents (e.g., their willingness to participate, the level of effort given), and other paradata related to survey features that may affect burden (e.g., length of interview, frequency and type of contact, doorstep concerns). We use structural equation modeling (SEM) to examine the associations between the above measures and latent factors that we hypothesize should impact burden, and then examine the nature of the relationship between these factors and respondents' perceptions of burden. We believe that this type of approach can significantly improve our understanding of burden by identifying characteristics of respondents that are most associated with high levels of burden given a particular survey feature (e.g., length). The ability to predict which respondents are at a greater danger of feeling burdened will help survey organizations to modify their survey protocols so as to reduce the likelihood of particular respondents feeling burdened and to reduce the negative impact of burden on data quality.

2. Methods

2.1 Data

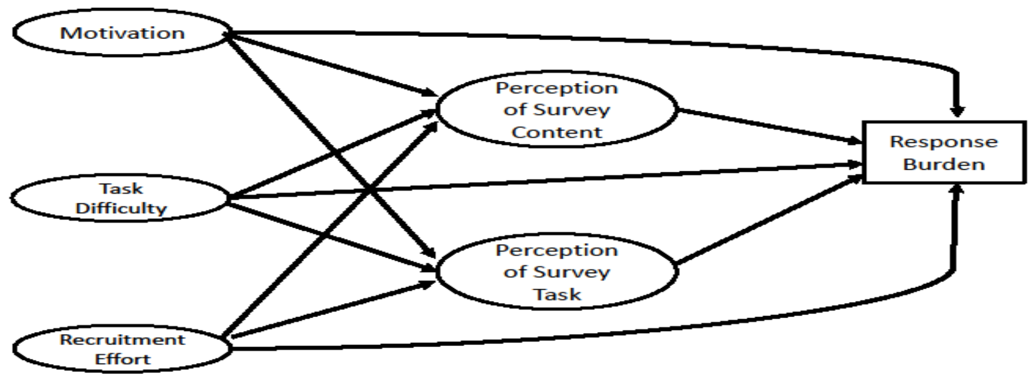
Data used for this study come from the Consumer Expenditure Quarterly Interview Survey (CEQ). The CEQ is a longitudinal survey sponsored by the Bureau of Labor Statistics (BLS). It collects comprehensive information on a wide range of consumers' expenditures and incomes, as well as the characteristics of those consumers. It adopts a rotation panel design and sampled households are interviewed five times before retiring from the panel.

Analyses are limited to panel members who completed their 5^{th} (and last) interview between October 2012 and March 2013, leading to a total of 3,340 cases for the analyses. At the end of their regular interview, these panel members were asked additional questions about their perceptions of and attitudes about CEQ in particular and surveys in general. One of the items directly asked how burdensome the survey was to them and the response categories are "not at all burdensome," "a little burdensome," "somewhat burdensome," and "very burdensome." This question makes the CEQ the only survey that asks directly about respondents' feeling of burden, unlike most of the earlier studies that measure response burden indirectly.

Additional data were extracted from the survey's Contact History Instrument (CHI). CEQ interviewers record information about each attempt to contact the sampled household, completing a brief set of standardized questions about the contact (e.g., mode, outcome) and, if contact is made, their assessment of the respondent, and any doorstep concerns expressed by the respondent (e.g., "not interested in the survey," "hostile behavior," "too busy," etc.). Information about the duration of the interview is automatically collected and also is stored in CHI.

2.2 Analytic Method

The data were analyzed using structural equation modeling (SEM). We selected this method because it allowed us to test a model of burden that includes latent factors related to survey features, respondent perceptions, and other respondent characteristics, and to examine the causal relations (direct and indirect) between these factors and burden. We used the SAS procedure PROC CALIS to estimate the model using a maximum likelihood estimation method.



1

Figure 1: Structural Model of Respondent Burden

2.3 Burden Model

The SEM framework consists of two inter-related models: (1) the measurement model, which describes the assignment of the observed items (or indicators) to each unobserved latent factor; and (2) the structural model, which describes the relationship among the set of latent factors. Both models are explicitly defined by the analyst, and depicted in a path diagram. For our structural model (see Figure 1), we examined two factors related to survey and task characteristics (*task difficulty* and *recruitment effort*), two factors related to respondent perceptions (*perceptions of the survey task* and *perceptions of survey content*), and one factor related to another key respondent characteristic, *motivation*. A description of the items used to construct the latent factors included in the structural model follows.

2.3.1 Measure of burden

Burden was measured through a dummy variable created from the survey item asking respondents directly how burdensome they felt the survey was, where 1 represents "very burdensome," or "somewhat burdensome" and 0 "a little burdensome" or "not at all burdensome." In the SEM, burden was treated as an observed variable with no measurement error. (We discuss this issue further in the Discussion section of the paper.)

2.3.2 Measure of motivation

Motivation in our model was measured as a latent construct with five indicators that are conceptually related to motivation. The first indicator, *cooperativeness*, represents whether the respondent was rated by interviewers as "very cooperative" or not.[1] The second indicator, *effort*, represents whether the respondent was rated by interviewers as putting in "a lot of effort" or not. The third indicator, *low concern*, was a summary variable drawing from doorstep concerns data and represents whether or not respondents have low or no concerns towards participation in the CE. The forth indicator, *trust*, represents whether or not respondents trusted in the US Census Bureau to safeguard the information they

[1] We ran SEM with indicators on their original scales. The models had a very poor fit. Instead, we dichotomized every indictor and reported results below using dichotomized indicators.

provided. The last indicator, *non-sensitivity of CE*, represents whether or not respondents considered CE as "not at all sensitive."

2.3.3 Measure of task difficulty

Task difficulty was measured as a latent construct with two indicators.[2] The first indicator, *long interview*, denotes whether or not the length of the interview was longer than the median length. The second indicator, *multi-person household*, represents whether or not respondents resided in a multi-person household. We selected this indicator because respondents living in multi-person households likely have more expenditures to report than those in single-only households, and because proxy reporting for other household members' expenditures may be more difficult (i.e., burdensome) than simply reporting for oneself.

2.3.4 Measure of recruitment effort

Recruitment effort was measured as a latent construct with three indicators. The first indicator, *frequent contact*, represents whether or not the respondent was contacted more frequently than the median number of contact attempts. The second indicator, *frequent personal visit*, represents whether or not respondents were visited by interviewers more frequently than the median number of personal visits. The third indicator, *converted refusal*, represents whether or not respondents went through refusal conversion process at least once.

2.3.5 Measure of perception of survey content

Perception of survey content was measured as a latent construct with two indicators. The first indicator, *interesting survey*, looks at whether or not respondents considered the CE as "very interesting" or "somewhat interesting." The second indicator, *easy survey*, represents whether or not respondents considered questions in the CE as "very easy" or "somewhat easy" to answer.

2.3.6 Measure of perception of survey task

Perception of survey task is measured as a latent construct with two indicators. The first indicator, *long survey*, represents whether or not the respondent considered the CE interview as "very long." The second indicator, *too many rounds*, represents whether or not respondents considered the five rounds of CE interviews as "too many."

3. Results

3.1 Model Fit Statistics

We examined several model fit statistics. The Chi-square test indicated a poor fit with the data ($X^2(77)=969$, $p<.0001$), though this measure can be an overly sensitive test of global fit with large sample sizes as we have in our study (Byrne 1998; Kline 1998). The second index of overall fit, the Standardized Root Mean Square Residual (SRMSR), was 0.056. According to O'Rourke and Hatcher (2013), SRMSR values less than 0.055 suggests a good fit and values less than 0.09 are suggestive of fair or adequate fit. Therefore, our SRMSR value indicates adequate to good model fit.

---

[2] Age and education are commonly used predictors of cognitive ability and difficulty (see Krosnick, 1991). We examined but ultimately dropped age and education as indicators for this construct in our final model due to poor fit.

We also looked at two parsimony indices. The Root Mean Square Error of Approximation (RMSEA) was 0.059, indicating a fair or adequate fit (O'Rourke and Hatcher, 2013).[3] The Adjusted Goodness of Fit Index (AGFI) was larger than 0.90, reflecting a good fit (AGFI=0.938). The Bentler Comparative Fit Index, an incremental index, was 0.868. This falls below the traditional 0.90 cut-off value, suggesting a poor fit.

Looking across all indices, we considered our models to reflect a fair fit to the data.

3.2 Measurement Model

Estimates from our SEM measurement model are shown in Table 1. The unstandardized factor loadings for each item with its associated latent variable are statistically significant and the standardized loadings are generally sizeable (i.e., greater than 0.3). All of the loadings are in the expected direction. For example, higher respondent motivation was associated with cooperative and effortful respondents who are trusting of the survey organization, express few doorstep concerns, and who do not view the survey questions as sensitive. *Recruitment effort* was positively associated with frequent contact attempts and personal visit contacts; the association between converted refusals and recruitment effort is weaker, but remains in the expected direction. Multi-person households and to a lesser extent longer than median interview durations were positively related to *task difficulty*. Negative *perceptions of the survey content* were reflected in respondents reporting less interest in and greater difficulty with the survey. Negative assessments of the survey length and the number of interviews were associated with negative *perceptions of the survey task*.

Table 1: Model Estimates for the Measurement Model

			Unstandardized		
Measurement Model		Standardized Estimates	Estimates	S. E.	*p*-value
<u>Factor</u>	<u>Indicator</u>				
Motivation	Cooperativeness	0.630	1		
	Effort	0.340	0.590	0.037	<0.0001
	Low concern	0.342	0.624	0.039	<0.0001
	Trust	0.366	0.563	0.033	<0.0001
	Non-sensitivity	0.370	0.608	0.035	<0.0001
	Frequent contact	0.801	1		
Recruitment Effort	Frequent PVs	0.600	0.745	0.055	<0.0001
	Converted refusal	0.295	0.324	0.028	<0.0001
Task Difficulty	Multi-person HH	0.721	0.349	0.069	<0.001
	Long interview	0.271	1		
Perception of Survey Content	Interesting survey	0.567	1		
	Easy survey	0.354	0.520	0.034	<0.0001
Perception of Survey Task	Long survey	0.670	0.919	0.033	<0.0001
	Too many rounds	0.638	1		

[3] RMSEA values less than .055 indicates a good fit and values less than .09 suggests a fair and adequate fit (O'Rourke and Hatcher, 2013)

3.3 Structural Model

With our measurement model validated, we examined the hypothesized structure of our latent factors. We begin by looking at the direct effects of each factor on the other model factors (see Table 2). The most striking result in this table is that the only factor in our model that had a significant direct effect on respondent burden was respondents' *perception of survey task*. Positive impressions of the survey task (reflecting respondents' subjective assessments that the length of the interview and frequency of the survey were reasonable) were associated with significantly lower levels of burden. However, the *task difficulty* factor itself, which in part reflects the effects of actual survey length, showed no direct statistical association with burden perceptions, and neither did *perception of survey content*.

Table 2: Model Estimates from the Structural Model of Burden

Structural Model		Standardized Estimates	Unstandardized Estimates	S. E.	*p*-value
Factor	Effect on				
Motivation	Perception of Survey Content	1.043	1.088	0.050	<0.0001
	Perception of Survey Task	0.999	1.231	0.057	<0.0001
	Burden	-0.712	-1.279	2.601	n.s.
Recruitment Effort	Perception of Survey Content	0.099	0.070	0.027	<0.05
	Perception of Survey Task	0.093	0.077	0.028	<0.01
	Burden	-0.048	-0.058	0.177	n.s.
Task Difficulty	Perception of Survey Content	-0.232	-0.182	0.049	<0.001
	Perception of Survey Task	-0.291	-0.270	0.064	<0.0001
	Burden	0.177	0.238	0.455	n.s.
Perception of Survey Content	Burden	0.767	1.323	2.289	n.s.
Perception of Survey Task	Burden	-0.698	-1.019	0.278	<0.001

The remaining effects shown in Table 2 generally are in the expected direction. For example, more highly motivated respondents were more likely to have a positive impression of the survey content and task than those who were less motivated. Similarly, greater task difficulty was associated with more negative perceptions of the survey content and task. The one puzzling finding is that greater recruitment effort (more contacts, more personal-visit contacts, and the need to carry out a refusal conversion) was associated with more *positive* feelings about both the survey task and the survey content. One possible explanation for this finding is that individuals who are administered more personal visit interviews also are more likely to be administered the post-survey burden questions in person, and therefore may be less willing to express negative reactions about their survey experience to their interviewer. Whatever the reason for the direction of the effects, the size of these effects is very small.

We next estimated the indirect and total effects of our model factors. Indirect effects are mediated by at least one intervening variable; total effects are equal to the sum of the direct and indirect effects. Table 3 summarizes the results of this decomposition of effects.

As shown in Table 3, *motivation* and *perception of survey task* had significant overall negative impacts on burden, and *task difficulty* had a significant overall positive impact on burden. Contrary to views commonly held in the survey field, the usual-suspect causes of burden such as *motivation*, *task difficulty*, and *recruitment effort* had no direct effects on burden. The indirect effects of these three factors also were not statistically significant, but this is due to the fact that *perception of survey content* had a positive impact on burden but the *perception of survey task* had a negative impact. As a result, the sum of these two paths essentially cancelled out the associated indirect effects. For instance, the indirect effect of motivation on burden through *perception of survey content* was 0.800 (i.e., 1.043*0.767=0.800). By contrast, the indirect effect of motivation on burden through *perception of survey task* was -0.697 (i.e., 0.999*-0.698=-0.697). The sum of the indirect effects for *motivation* then was 0.103 (=0.800-0.697), which was not significant.

Table 3: Decomposition of Effects of Latent Factors on Burden

	Total Effects	Direct Effects	Indirect Effects
Motivation	-0.609***	-0.712	0.103
Recruitment Effort	-0.036	-0.048	0.012
Task Difficulty	0.202***	0.176	0.026
Perception of Survey Content	0.767	0.767	0
Perception of Survey task	-0.698***	-0.698**	0

*p<0.05;**p<0.01; ***p<0.001

4. Discussion

Prior research on response burden often has relied on inadequate conceptualizations and measures of burden. It is not surprising then that it is difficult, based on the small empirical literature that exists, to make firm predictions about survey features or respondent characteristics that are most likely to give rise to burden. In this study we developed and tested a model that assumes that burden is a subjective phenomenon, affected by respondents' psychological responses to various elements of the survey. The survey data we used to test this model included direct measures of those reactions from respondents themselves, interviewer assessments, and objective features of the survey. We used structural equation modeling to assess how well these data fit latent factors we hypothesized to be important contributors to perceived burden, and then evaluated the impact those factors had on burden.

The results of this study validated our underlying measurement model – our indicators were all significantly related to their associated latent variables in the expected direction. Results of our structural model showed that respondents' subjective perceptions of the survey task (e.g., reactions to the interview length and difficulty) had a significant direct impact on burden, but the objective survey features themselves did not. Respondent motivation and survey features thought to affect task difficulty (e.g., interview length, proxy reporting) did have significant overall effects on burden, but only because we modeled their impact on respondents' perceptions of the survey task and content.

These findings support the notion of burden as a subjective, multidimensional phenomenon, and underscore that the impact of any given survey feature on burden will vary across respondents on the basis of measurable subjective reactions and attitudes. Our analysis was made possible because of the CEQ's unique datasets which contain information about respondent reactions to the survey, and we hope our results encourage other surveys to collect and disseminate similar data. Continued work should be done to test and refine measures of respondents' reactions and their perceptions about survey features as well as other domains that could impact perceived burden (e.g., privacy attitudes, civic engagement, etc.). Incorporating these measures will result in more robust and reliable models of burden, and allow researchers to more effectively plan and test targeted interventions designed to reduce respondent burden, and better examine associations between burden and data quality.

A key limitation of this study was that there was only one survey question asking directly about the feeling of burden. As a result, the SEM treated this observed indicator as error-free. Of course, this is a strong assumption, and almost certainly violated. We used an alternative approach suggested in Kline (1998: p264-266) in which we re-specified our model by treating the observed burden variable as an observed indicator of a latent burden factor. We used the Survey Quality Predictor program (http://sqp.upf.edu/) to obtain an estimate of the quality and the error of the burden item. We then reran the SEM using the error estimates from the SQP to fix the measurement error term of the observed burden indicator and fixing the loading of the burden indicator on the latent burden construct to be 1. The conclusions remained unchanged. However, the SQP estimate of the error term is at best only an approximation of the true error.

We also were limited to the number and type of indicator variables available on the CEQ files. Three latent factors were measured with two indicators, and this is not ideal (see Kenny, 1979). We believe that the model fit might be improved if more indicators were available and could be included in the measurement model. Similarly, our structural model might be improved by introducing interaction terms between our three exogenous latent factors. It is possible that the interaction of *motivation* and *task difficulty* could have a direct impact on burden even though *motivation* and *task difficulty* alone did not have a significant direct effect. Including the interaction terms also may improve the model fit.

Acknowledgements

This research was funded by the ASA/NSF/BLS Research Fellowship awarded to Ting Yan.

References

Bradburn, N. (1978). Respondent Burden. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 1978: 35-40.

Byrne, B. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming.* Mahwah, NJ: Lawrence Erlbaum Associates.

Fricker, S., Gonzalez, J., & Tan, L. (2011). Are You Burdened? Let's Find Out. *Paper Presented at the Annual Conference of the American Association for Public Opinion Research*, Phoenix, AZ.

Fricker, S., Kreisler, C. and Tan, L. (2012). "An exploration of the application of PLS path modeling approach to creating a summary index of respondent burden." *In JSM Proceedings*, Survey Methods Research Section, Alexandria, VA: American Statistical Association, 4141-4155.

Haraldsen, G. (2004) Identifying and Reducing Response Burden in Internet Business Surveys. *Journal of Official Statistics*, 20, 393-410.

Hedlin, D., Dale, T., Haraldsen, G., and Jones, J. (2005). *Developing Methods for Assessing Perceived Response Burden*. Research report, Stockholm: Statistics Sweden, Oslo: Statistics Norway, and London: Office for National Statistics.

Kline, R. B. (1998). *Structural Equation Modeling*. New York: Guilford Press.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213–236

O'Rourke, Norm and Hatcher, Larry (2013). A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling. SAS Institutes.

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