Interviewer Effect on Conversational Interviews: Results from the 2010 & 2011 SIPP-EHC

Rachael Walsh, PhD
rachael.walsh@census.gov
U.S. Census Bureau; Social, Economic, and Housing Statistics Division

Population Association of America 2012 Annual Conference
Session 1003—Innovations in Data, Methods, and Measure

Disclaimer: The views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.[1]

[1] All comparative statements in this report have undergone statistical testing, and, unless otherwise noted, all comparisons are statistically significant at the 95% significance level.

Abstract: The new conversational interviewing style associated with an Event History Calendar is a departure from the conventional scripted style of Census Bureau surveys. The new Survey of Income and Program Participation include an Event History Calendar (SIPP-EHC) and is currently undergoing annual field tests. The role of the interviewer and differences in interview administration by interviewer is of particular interest. Between 2010 and 2011, the interview length decreased by 17 minutes, but this is not solely attributable to the interviewers because the instrument is still in development. This research uses multilevel modeling to assess the interviewer effect on individual interview length using data from the 2010 and 2011 SIPP-EHC field tests. Interviewer effect decreased 3% from 2010 to 2011. In 2011, 6% of the variation in interview length was attributable to interviewer effect. This makes the interviewer effect on the conversational 2011 SIPP-EHC comparable to the scripted 2008 SIPP interviews.

The importance of the interviewer in collecting high quality data during face-to-face interviews has been well-documented (Campanelli and O'Muircheartaigh, 2002; Couper and Groves, 1992; Groves and Couper, 1998; Japec and Lundqvist, 1999; Loosveldt *et al*., 2002; Martin and Beerten, 1999; Nicoletti and Buck, 2003). The effect of the interviewer on the length of the interview, however, has not produced a cohesive body of literature. Evidence suggests respondents who are engaged and enjoy the interview experience have longer interviews (Branden *et al*., 1995; Hill and Willis, 1998; Zabel, 1998). However, shorter interviews have been found to result in both higher response rates and lower sample attrition in longitudinal surveys (Bogen, 1996; Zabel, 1998). The Office of Management and Budget (OMB) requires survey questionnaires collect information in a manner that has been cognitively tested, is relevant and non-duplicative, and the questionnaire be of reasonable length.[2]

Survey designers must then create instruments that contain the critical subject matter that is asked in a clear, concise manner, while ensuring this material is relevant to respondents and the survey itself is not too long. Interviewers have to be able to use this survey instrument in the field and engage respondents in a meaningful way. Personal interviews typically provide interviewers with a script to follow when asking respondents for information to ensure data continuity (Groves, 1989; Schuman and Presser 1981; Sudman, Bradburn, and Schwarz 1996). Conversational interviews, however, can enhance rapport with the respondent, something crucial for longitudinal surveys when attempting to minimize sample attrition (Lavin and Maynard, 2001). It may be possible for a hybrid instrument—one that combines both scripted and

---

[2] OMB Guideline 2.3.1 from *Standards and Guidelines for Statistical Surveys,* http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf .

conversational interviewing into one survey—to aid recall and build this rapport with respondents, while still ensuring the same information is collected across interviewers.

This research begins assessing this hybrid survey, measuring the interviewer effect based on one measure of respondent burden, interview length. As the current body of literature does not equate shorter interviewers with better interviews, this research does not focus on the interview length per se, but merely relies on this measure to determine the interviewer effect on interview length in a survey that has a conversational element imbedded in a scripted interview. To do so, this research addresses the following research questions:

1. How much variation in conversational interview length can be attributed to the interviewer effect;
2. How much variation in conversational interview length can be attributed to regional effects;
3. Are these effects decreasing as both interviewers and regions gain more practice fielding the new conversational style of interview?

To answer these research questions, the redesign of the Survey of Income and Program Participation (SIPP) can be helpful. The production SIPP is a scripted interview, and as part of the redesign process, a conversational event history calendar (SIPP-EHC) has been integrated as a central part of the survey instrument. The SIPP-EHC will take the place of the production SIPP in 2014.

Redesigning the SIPP

The SIPP is a longitudinal, nationally representative survey of U.S. households. At the beginning of each panel, 30,000 to 65,000 households are sampled and interviewed three times per year for a period of three to four years. Once an individual is interviewed in the first wave, the SIPP then follows that person and interviews any additional people living with that original sample person for the duration of the panel. The overall purpose of the SIPP is to collect information regarding all sources of income to demonstrate the dynamics of the U.S. economy and how the economic situation in households change over time across the country. In addition to capturing income from a variety of sources including participation in social welfare and social insurance programs, the SIPP also provides information with respect to the well-being of children and adults, childcare arrangements, health care coverage, and a variety of other areas affecting the financial and social situation of U.S. families and households.

In an attempt to reduce both the financial and respondent burdens of fielding the production SIPP three times a year, the U.S. Census Bureau is incorporating a new method of conversational annual data collection—the Event History Calendar (EHC) into a reengineered SIPP data collection instrument. EHCs enable the respondent to recall events over an extended period of time through the use of memory anchors (Callegaro, Belli, Serrano, and Palmer, 2007). The reengineering of the SIPP combines collection of data representing the traditional SIPP core interview with data elements from most of the existing Topical Modules using an EHC, in a new product called the SIPP-EHC. The use of a full calendar year has been found to be the most beneficial form of EHC data collection for the respondent, which allows the SIPP-EHC to collect the same high quality data in one annual visit instead of three visits (once every four months), additionally reducing seam bias (Callegaro and Belli, 2007).

The SIPP-EHC collects most of the information found in the core of the SIPP through the same, scripted methods as the SIPP, using similar, and in many cases, identical question wording. However, the calendar portion of the interview—which is also the conversational portion of the instrument—collects the information with respect to residency, marital history, employment, means tested social welfare program participation, and health insurance coverage. These topics are often interdependent making the contextual grouping of these topics an easy source of conversation for interviewers who are more accustomed to scripted interviews. As such, the SIPP-EHC provides a platform for testing the proposed research questions, as well as a potential instrument design for longitudinal data collection that can satisfy OMB requirements while building the rapport necessary to minimize attrition and improve recall.

The EHC relies on memory cuing to aid respondents in reporting accurate information across the longer reference period in the redesigned SIPP-EHC (Belli, 1998). Specifically, the EHC utilizes top-down, sequential, and parallel cuing. Top-down cuing relies on the preferential treatment of high priority items in the memory as starting points, allowing respondents to access the most memorable events and work down through the less memorable events (Barsalou, 1988; Conway, 1996). The SIPP-EHC introduces the calendar portion of the interview by collecting landmark events—or memories of import to the respondent that can be used later in the calendar as memory anchors—utilizing top-down cuing.

In sequential cuing, chronological ordering of events is relied upon to aid the respondent's memory recall within a topic (Belli, Shay, and Stafford, 2001). The conversational style of the calendar allows respondents to report life events in a manner conducive to his/her specific memory strengths, capitalizing on the type of memory cuing with which the respondent most readily identifies. In other words, respondents are able to narrate the things that happened

during the reference period (Brown and Schopflocher, 1998; Schank and Abelson, 1995). Once the landmarks are recorded, the flexibility of the calendar permits the sequential reporting of events while enabling interviewers to use parallel cuing to further enhance recall (Conrad and Schober, 2000).

Parallel cuing uses related events across topics to aid respondents in recalling specific information (Belli, Shay, and Stafford, 2001). To summarize, the EHC interviewing style first prompts respondents to recall the dates of the most memorable events, which provoke the memory of other, related events, while chronology is then used to fill-in the gaps. The inclusion of an EHC in the SIPP is a significant advance that enables the expansion of the reference period from four to twelve months, while still collecting the same quality data. However, this conversational style of interviewing is new to the U.S. Census Bureau. The SIPP-EHC project can be used to measure the interviewer and regional effects on the length of conversational interviews as a measure of the improvements made to the SIPP-EHC between fielding iterations.

The SIPP-EHC Project

The SIPP-EHC has undergone three field tests to date. The first version of the SIPP-EHC was a paper and pencil interview fielded in 2008. The 2010 SIPP-EHC test was the first automated version of the reengineered SIPP to use Blaise software and a calendar designed using C# software. As this was the first version of any SIPP interview to utilize these software packages, the instrument had a few functionality issues. These "bugs" made interviewer training difficult and resulted in an overly complicated interview. This contributed to increased respondent burden through lengthy interviews.

While it has been noted that interview length is a complicated measure of respondent burden (Bogen, 1996; Branden *et al*., 1995; Hill and Willis, 1998; Zabel, 1998), in the specific case of the 2010 SIPP-EHC, we know interview length was an issue. As the SIPP-EHC is still being tested, interviewers are debriefed after the interviewing period, one of the ways this information has been obtained. Key modifications were made to the instrument for the 2011 field test that significantly impacted the length of the interview. Many of the instrument bugs were identified and rectified. Additionally, an income screener was added to the instrument preventing those with disqualifying levels of income from being asked many questions about means tested social welfare programs.

The length of the EHC reference period was extended from the 12 month calendar year to the calendar year plus the months in the interview year up to and including the interview month. To enhance memory recall, the data collection instrument was reformatted to ask respondents about his/her current situation, then moving backwards in time through the reference period. This time sequencing of retrospective reporting has been found to improve data quality and reduce interview length (Belli, 1998).

After the completion of all individual interviews in the household, both the respondent and the interviewers were asked to provide feedback regarding the interview, and these comments were used to make additional improvements to the instrument. One specific change was to copy residency information provided by the first respondent to all respondents identified as living together during the entire reference period. Additionally, marital history information is copied to the married partner's record. The copying of previously collected data from one person's interview to other(s) reduces the number of questions asked of subsequent respondents.

In addition to the improvements made to the instrument between 2010 and 2011, the sampling frame, training, and field administration of the survey also underwent changes.

Data & Methods[3]

In 2010, the SIPP-EHC was conducted in only six of the twelve regional offices, with an initial sample size of 7,982 households, and an 81.9% household response rate. For the SIPP, household members age 15 and over are treated as adults and are interviewed individually. In the SIPP-EHC, children do have records, and even their own calendars, however, this information is provided via proxy by a knowledgeable household member. The burden of child interviews is on the person providing the proxy, not the child. Additionally, child interviews are substantially shorter than adult interviews, with the majority of the calendar (i.e. marital history and employment) being omitted. Therefore, child interviews were not included. In this research, 11,058 adult interviews from the 2010 survey are included, serving as the 2010 SIPP-EHC sample for this analysis.

In 2011, the survey was conducted in all twelve regional offices, but with an overall sample size reduced to 4,051 households. The household response rate improved from 2010 to 84.9% in 2011. Child interviews were again not used, but interviews were conducted with 4,864 adults. The sample for the 2010 and 2011 SIPP-EHC field tests oversamples low-income Census tracts. Regions then receive cases and make assignments to interviewers usually based on spatial proximity. Multilevel modeling takes into consideration the clustering that occurs as a result of

[3] The SIPP-EHC is still in the testing phase—meaning the sample is not nationally representative, and the data are subject to error.

both the sampling, as well as the actual fielding, of the survey (Campanelli and O'Muircheartaigh 2002; Hox 1994; Olson and Peytchev 2007).

Audit trail files were used to provide the outcome of interest—*Interview Length*. Audit trail files capture every keystroke made by an interviewer during a case, providing each entry with a date and time stamp. This provides detail with respect to interview length that allows both individual and household level interview length to be calculated. Interview length was then available for all households where at least one individual was interviewed.

Table 1 displays the descriptive statistics for both the 2010 and 2011 samples. The table displays the individual and interviewer characteristics used in the models, as well as the sample sizes. The 2010 sample size consisted of 8,154 adult interviews with an average of 31 adults per interviewer. The 2011 sample for this research consisted of 3,821 adult interviews with an average of 22 adults per interviewer.

The interpretation of intercepts is more meaningful in multilevel modeling if they are dichotomously recoded or mean centered (Raudenbush and Bryk 2002). In these models, the respondent's age and education levels were used because both age and education can have an effect on respondents' interactions with interviewers (Davis and Silver 2003; Olson and Peytchev 2007). *Age* and *Education* were both centered to the sample mean. Both employment and participation in social welfare programs require additional interview questions, increasing the interview length. Income is often a disqualifier for social welfare programs. *Unemployed* thus receives the positive value in the dichotomous recode. Those who participate in any of the following means tested social welfare programs—Food Stamps or Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), General

Assistance, or Women, Infants, and Children (WIC)—are categorized together as *Social Welfare Program Recipients*.

Additionally, the language spoken in the home has a significant effect on the length of the interview (Walsh, 2012). The household respondent is asked whether another language is spoken in the home, and then asked to identify who speaks another language and to identify the language. *ESL* dichotomously identifies those who speak a language other than English in the home. The intercept can be interpreted as the average interview length for a respondent of average age and level of education, who is employed, who is not receiving social welfare programs, and who only speaks English (Raudenbush and Bryk 2002). The first model also includes the natural log of *Interview Order* to account for interviewer improvement during the interviewing period (Olson and Peytchev, 2007). The equation follows:

Level 1: Respondent Model

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}\,\mathrm{Ln}(\mathrm{Order}_{ijk}) + \beta_{2jk}\,(\mathrm{Age} - \overline{\mathrm{Age}}\,)_{2jk} + \beta_{3jk}\,(\mathrm{Educ} - \overline{\mathrm{Educ}}\,)_{3jk} + \beta_{4jk}\,\mathrm{Employment}_{4jk} + \beta_{5jk}\,\mathrm{Program\ Participation}_{5jk} + \beta_{6jk}\,\mathrm{ESL}_{6jk} + r_{ijk}$$

The second level of the model introduces interviewer characteristics[4]. Though each region is permitted autonomy when assigning cases, in most cases, supervisory interviewers are reserved until later in the interviewing period, and assigned the more difficult cases, specifically refusal conversions. *Supervisory Status* was included as a dichotomous variable such and the

[4] In accordance with the U.S. Census Bureau policy, the interviewer demographic characteristics cannot be used in analyses in order to protect interviewer privacy. A new policy is being reviewed to make these characteristics available under certain circumstances in the future.

intercept is reflective of non-supervisory status. *SIPP Experience*, and in 2011, *SIPP-EHC Experience*, were also dichotomously coded, with experience receiving a positive value. Years of *Census Experience* was mean centered, as was *Certification Test Score* which the interviewer received on the certification exam given at the end of the classroom training and prior to fielding the survey. The intercept is the average length of an interview conducted by a non-supervisory interviewer, with neither SIPP nor SIPP-EHC experience, in relation to the average length of Census Bureau service and average exam score, and was determined using the equations below:

Level 2: Interviewer Model

$$\beta_{0jk} = \Upsilon_{000} + \Upsilon_{010}\,\text{Supervisory Status}_{010} + \Upsilon_{020}\,\text{SIPP Experience}_{020} + \Upsilon_{030}\,\text{SIPP-EHC Experience}_{030} + \Upsilon_{040}\,(\text{Census Experience} - \overline{\text{Census Experience}})_{040} + \Upsilon_{050}\,(\text{Certification Test Score} - \overline{\text{Certification Test Score}})_{050} + \mu_{pjk}$$

$$\beta_{1jk} = \Upsilon_{100} + \Upsilon_{110}\,\text{Supervisory Status}_{110} + \Upsilon_{120}\,\text{SIPP Experience}_{120} + \Upsilon_{130}\,\text{SIPP-EHC Eperience}_{130} + \Upsilon_{140}\,(\text{Census Experience} - \overline{\text{Census Experience}})_{140} + \Upsilon_{150}\,(\text{Certification Test Score} - \overline{\text{Certification Test Score}})_{150} + \mu_{pjk}$$

The third and final level of the model was at the regional level. All respondent and interviewer characteristics are fixed, allowing the means to vary by interviewer and region alone, as shown in the following equation:

Level 3: Regional Model

$$\beta_{2jk} = \Upsilon_{200} + \mu_{10k}$$

In addition to the outcomes from the multilevel modeling, the intra-class correlation coefficient (ICC) was also calculated at each level to provide the variance attributable to interviewer effects in each phase of the modeling, and regional effects in the final model (Snijders and Bosker, 1999).

Results[5]

The results from the multilevel models are presented in Table 2. The average interview length for employed respondents whose primary language is English and who are not participating in social welfare programs was 58.3 minutes in 2010. In 2011, the interview length for the same type of respondent was 41.5 minutes. While interviewers had a steep learning curve in 2010, this level was reduced in 2011 as seen in Figure 1. In 2010, by the 10th interview, interviewers were able to reduce the interview length by approximately 10 minutes. In 2011, however, the reduction in interview length by the 10th interview was approximately 5 minutes. While both curves have steeper slopes initially, in 2010 the slope continued to fall while in 2011 the slope begins to flatten. By the 25th interview, interview length was reduced by 15 minutes in total in 2010, but only by 7 minutes in 2011.

All of the respondent characteristics were statistically significant contributing factors in this analysis (see Table 2). The effects of employment and social welfare receipt, while still statistically significant, were less in 2011 than 2010. The effect of the language barrier for non-English speakers in the more conversational interviewing style was not statistically significant in

[5] This analysis was done without weights and does not incorporate the design effect. The results are not generalizable to the population and my results are particular to this specific sample. The two contributing factors are the oversample of low income and the presence of nonresponse.

2010 or 2011. In 2010 supervisory interviewers had significantly shorter interviews than non-supervisory interviewers by 7.1 minutes. While this effect was not seen in 2011, having SIPP interviewing experience did decrease the 2011 interview length by 5.0 minutes. Both SIPP-EHC experience and the score on the interviewer certification exam were statistically significant predictors of interviewer length in Model 2 in 2011; however, when regional effects were added to the model in Model 3, these covariates were no longer statistically significant.

Looking at the ICC reported for Model 3 in Table 2, in 2010, 10% of the variation in interviewer length was attributable to interviewer effects. An additional 1% of the variation was attributable to the regional effect. In 2011, 6% of the variation in interview length was attributable to the interviewer effect, and 1% could be attributed to regional differences. These results are encouraging and indicate improvements to not only the survey structure, but also the functionality of the data collection instrument as discussed in the next section.

Discussion & Conclusions

The instrument enhancements from 2010 to 2011 resulted in both a reduction in respondent burden as well as a reduction in both interviewer and regional effects. While the interviewer effect decreased by 3%, the regional effect was reduced to almost half. Additionally, the 2011 SIPP-EHC interviewer and regional effects on interview length are much more comparable to the interviewer effects seen in production SIPP (author, under review). Improvements continue to be made to both the instrument and the interviewer training in hopes to further reduce respondent burden and interviewer effects.

The survey is currently undergoing redesign and, both the instrument and the interviewer training are continuously improving. Therefore, the gains between the 2010 and 2011 tests cannot be attributed solely to enhanced EHC interviewing techniques. There have been multiple instrument improvements such as the introduction of an income screener, resolution of earlier navigation "bugs" with the instrument, copying of respondent answers from one household member to another, and more cohesive interviewer training, together contributing to the 17 minute reduction in adult interview length between the two tests. Since the average SIPP-EHC household has two adults, this is a savings of 34 minutes per household.

In the 2010 test interviewer training, emphasis was placed on the fact that it was a "test", not a production instrument. This emphasis led regional offices involved in the 2010 test to relax person non-response rate requirements. This was not the case in 2011 where person non-response rate requirements were similar to that of production interviewing. It is therefore likely that the 17 minute reduction in interview length between 2010 and 2011 is a conservative estimate. The interviewer effect on respondent burden is also affected by the regional office choices regarding the actual case assignment to specific interviewers.

The addition of the remaining six regional offices in the 2011 test also produced interesting results. Six of the twelve regional offices were involved in the 2010 SIPP-EHC field test. In the 2011 analysis, when the regional effects were modeled, they moderated for the effect of SIPP-EHC interviewing experience as well as the effect of the score on the certification exam. The certification exam and SIPP-EHC interviewing experience were both only done in half of the regional offices in 2010, which would indicate experience fielding the SIPP-EHC is beneficial. Both the SIPP-EHC and the certification test were done in only six of the regional offices in 2010. While neither interviewers nor regions can control respondent characteristics,

this research does provide some guidance for minimizing interviewer and regional effects on respondent burden. The level of SIPP-EHC experience will continue to increase with subsequent tests. The instrument, training, and the certification test will continue to be improved overtime.

The regional offices can focus on modifying current practices to maximize the benefits from the current level of experience. Current procedures require newly hired interviewers to be observed by his/her direct supervisor for at least the first two interviews. Newly hired interviewers may benefit more from observing a SIPP or SIPP-EHC experienced interviewer for the first interview, then have the experienced interviewer observe the new hire on the second interview. While the current supervisory interviewers are usually Census Bureau experienced, they are not always SIPP or SIPP-EHC experienced.

The instrument and training continue to be improved to minimize the respondent burden resulting from additional questions coming on path as a result of respondent characteristics or answers. The regions could further decrease respondent burden through a slight modification to the current practices of training and observing interviewers. Regardless of the improvements that the SIPP-EHC continues to undergo, the results of this analysis demonstrates that, despite the new, more conversational style of event history calendar interviewing, the interviewer and regional effects are decreasing to those of the production SIPP. The SIPP-EHC is demonstrating its ability to reduce costs and respondent burden compared to production SIPP.

While additional analysis is needed, this research demonstrates the effectiveness of using EHC interviews for data collection agencies. Conversational interviewing can enhance rapport and minimize attrition in longitudinal surveys (Boots-Miller, et al 1998; Hill and Willis 2001; Thornton, Freedman, and Camburn 1982). This evaluation demonstrates the feasibility of

incorporating an EHC. The SIPP-EHC has and continues to be evaluated and improved under the guidance of the U.S. Census Bureau.

References

Barsalow, L.W. (1988). The Content and Organization of Autobiographical Memories. In Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory, ed. Ulric Neisser and Eugene Winograd, pp. 193-243. New York: Cambridge University Press

Belli, R. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory*, 6, pp. 383-406.

Belli, R.F., Shay, W.L., and Stafford, F.P. (2001). Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly*, 65(1), 45-74.

Bogen, K. (1996), 'The Effect of Questionnaire Length on Response Rates – A Review of the Literature', *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pp. 1020-1025.

Boots-Miller, BJ, KM Ribisl, CT Mowbray, WS Davidson, MA Walton, SE Herman. (1998). Methods of Ensuringf High Follow-Up Rates: Lessons from a longitudinal study of dual-diagnosed participants. *Substance Use and Misuse.* 33(13):2665-85.

Branden, L., Gritz, R.M. and Pergamit, M. (1995), 'The Effect of Interview Length on Attrition in the National Longitudinal Study of Youth', National Longitudinal Surveys Discussion Paper no. 28, Bureau of Labor Statistics, US Department of Labor, Washington DC.

Brown, N.R., and Schopflocher, D. (1998). EventC lusters:A n Organizationo f Personal Events in Autobiographical Memory. *Psychological Science*, 9:470-75.

Callegaro, M. & Belli, R. (2007). Impact of the Event History Calendar on Seam Effects in the PSID: Lessons for SIPP. Paper presented at *The Use of Event History Calendar (EHC) Methods in Panel Surveys*, Washington, DC. December 5-6, 2007.

Callegaro, M., Belli, R., Serrano, E. & Palmer, D. (2007). Cultural variability in event history calendar and convention questionnaire interviews: A verbal behavior analysis. In *2006 Proceedings of the American Statistical Association,* 61st Annual conference of the American Association for Public Opinion Research, Alexandria, VA: American Statistical Association.

Campanelli, P. & O'Muircheartaigh, P. (2002). The Importance of Experimental Control in Testing the Impact of Interviewer Continuity on Panel Survey Nonresponse, *Quality & Quantity*, 33, 129-144.

Conway, M.A. (1996). Autobiographical Knowledge and Autobiographical Memories. In Remembering Our Past: Studies in Autobiographical Memory, ed. David C. Rubin, pp. 67-93. New York: Cambridge University Press.

Couper, M.P. and Groves, R.M. (1992), 'The Role of the Interviewer in Survey Participation', *Survey Methodology*, vol. 18, no. 2 (December), pp. 263-277.

Davis, D. & Silver, B. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science*, 47(1), 33-45.

Freedman, D., Thronton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: a technique for collecting retrospective data. *Sociological Methodology,* 18, 37-68.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Groves, R.M. and Cooper, M.P. (1998), *Nonresponse in Household Interview Surveys*, John Wiley and Sons, New York.

Hill, D.H. and Willis, R.J. (2001). Reducing Panel Attrition: A search for effective policy instruments. *The Journal of Human Resources*, 36(3), 416-438.

Huffcutt, A. I., & Woehr, D. J. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior, 20*(4), pp. 549-561.

Japec, L. and Lundqvist, P. (1999), 'Interviewer Strategies and Attitudes', Paper presented at the International Conference on Survey Non-Response, Portland, Oregon, 28-31 October.

Lavin, D. and Maynard, D.W. (2001). Standardization vs. Rapport: Respondent Laughter and Interviewer Reaction during Telephone Surveys. *American Sociological Review*, 66(3), 453-479.

Loosveldt, G., Pickery, J. and Billiet, J. (2002), 'Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey', *Journal of Official Statistics*, vol. 18, no. 4 (December), pp. 545-557.

Martin, J. and Beerten, R. (1999), 'The Effect of Interviewer Characteristics on Survey Response Rates', Paper presented at the International Conference on Survey Non-Response, Portland, Oregon, 28-31 October.

Nicoletti, C. and Buck, N. (2003), 'Explaining Contact and Refusals in the British and German Household Panels', in N. Buck, C. Nicoletti, A. McCulloch and J. Burton, *Report on Attrition Analysis and Item Non-response*, Chintex Working Paper no. 16, February.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A (Statistics in Society), 161*(1), pp. 63-77.

Olson, K. & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly*, 71(2) 273-286.

Raudenbush, S.W. and Bryk, A.S. (2002). Hierarchical Linear Models (Second Edition). Thousand Oaks: Sage Publications.

Schuman, Howard, and Stanley Presser. 1981. Questions and Answers in Attitude Surveys. San Diego: Academic Press.

Schank, Roger C., and Robert P. Abelson. 1995. "Knowledge and Memory: The Real Story." In Advances in Social Cognition (vol. 7), ed. Robert S. Wyer, pp. 1-85. Hillsdale, NJ: Erlbaum.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

Snijders, T.A.B and Bosker,R.J. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. Thousand Oaks: Sage Publications.

Thornton, A., Freedman, D., and Camburn, D. (1982). Obtaining Respondent Cooperation in Family Panel Studies. *Sociological Methods and Research*, 11:33-51.

Walsh, R. (2011). Certification Tests as an Indicator of Interviewer Productivity. *Population Association of America* Annual Conference, March 2011, Washington, DC.

Walsh, R. (2012). Language Barriers to Conversational Interviewing: Evaluating a new method of data collection. *American Association for Public Opinion Research* Annual Conference, May 2012, Orlando, FL.

Watson, N. and Wooden, M. (2009). Identifying Factors Affecting Longitudinal Survey Response. In Methodology of Longitudinal Surveys, Lynn, P. (Ed). John Wiley & Sons, Ltd.

Zabel, J.E. (1998), 'An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labour Market Behavior', *The Journal of Human Resources*, vol. 33, no. 2 (Spring), pp. 479-506.

Table 1. 2010 and 2011 SIPP-EHC Sample Descriptive Statistics.

	2010				2011			
	Interviewer	Supervisor	All		Interviewer	Supervisor	All	
Respondent Characteristics								
Age	41.01	42.60	41.08		42.10	42.32	42.11	
Education	2.02	2.05	2.02		2.18	2.15	2.18	
Unemployed	0.50	0.57	0.51	*	0.47	0.43	0.75	
Social Welfare Program Recipient	0.22	0.23	0.22		0.24	0.27	0.24	
English as a Second Language (ESL)	0.53	0.47	0.53	*	0.44	0.42	0.44	
Interviewer Characteristics								
Supervisory Status	--	--	0.14		--	--	0.19	
SIPP Experience	0.35	0.98	0.44	*	0.61	1.00	0.68	*
SIPP-EHC Experience	†	†	†		0.24	0.50	0.26	
Census Experience	2.28	11.61	3.56	*	2.84	10.26	3.38	*
Certification Test Score	0.76	0.78	0.76		0.77	0.84	0.77	*
Average # respondents/interviewer	34	11	31	*	25	6	21	*
Max # respondents/interviewer	153	49	--		106	21	--	
Respondents	10,549	509	11,058		4,606	258	4,864	
Interviewers	265	42	307		183	44	227	

Source: U.S. Census Bureau. Survey of Income and Program Participation-Event History Calendar (SIPP-EHC), 2010 and 2011.

* indicates statistically significant Wald Test at the 0.05 level.

† Not Applicable for the 2010 SIPP-EHC.

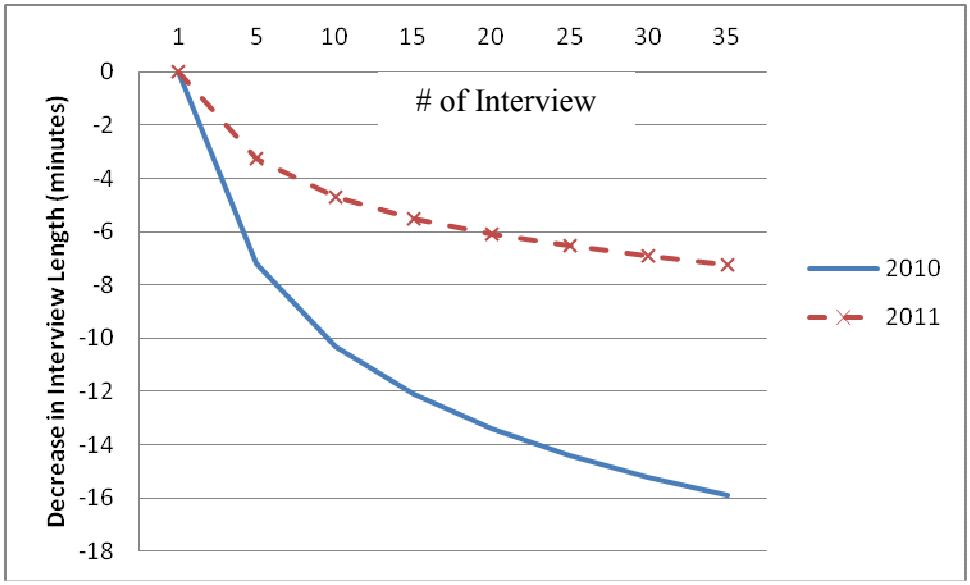
Table 2. Estimated Coefficients, Standard Errors, and Variance Components for Multilevel Models Predicting Interview Length in Minutes, 2010 and 2011 SIPP-EHC

	2010 SIPP-EHC			2011 SIPP-EHC		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Fixed Effects						
Respondent Level						
Intercept	56.52***	58.11***	58.33***	37.68***	41.45***	41.47***
	(0.89)	(1.28)	(1.82)	(1.17)	(1.54)	(1.66)
Ln(interview Order)	-4.55***	-4.48***	-4.47***	-2.12***	-2.01***	-2.03***
	(0.28)	(0.28)	(0.28)	(0.37)	(0.42)	(0.42)
Age	0.20***	0.21***	0.21***	0.24***	0.24***	0.24***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Education	1.86***	1.84***	1.88***	1.96***	1.90***	1.93***
	(0.22)	(0.27)	(0.27)	(0.32)	(0.36)	(0.36)
Unemployed	-15.89***	-15.46***	-15.45***	-7.56***	-7.34***	-7.29***
	(0.46)	(0.54)	(0.54)	(0.69)	(0.76)	(0.76)
Social Welfare Program Recipient	12.27***	12.16***	12.21***	8.66***	7.97***	8.03***
	(0.54)	(0.64)	(0.64)	(0.80)	(0.88)	(0.88)
ESL	2.57***	2.80***	2.79***	2.71***	2.67***	2.74***
	(0.49)	(0.58)	(0.58)	(0.75)	(0.82)	(0.82)
Interviewer Level						
Supervisory Status		-5.95*	-7.09**		-3.99	-4.07
		(2.67)	(2.59)		(3.36)	(3.35)
SIPP Experience		-2.33	-2.56		-4.88**	-5.01**
		(1.50)	(1.43)		(1.86)	(1.85)
SIPP-EHC Experience		†	†		-3.48*	-3.60
					(1.18)	(1.95)
Census Experience		-0.32	-0.28		0.13	0.18
		(0.20)	(0.19)		(0.20)	(0.20)
Certification Test Score		-2.11	-5.09		10.52*	8.80
		(5.67)	(5.55)		(5.14)	(5.21)

(**Table 2. Continued**)						
Random Effects						
Respondent Level	8.46	8.05	3.14	7.72	6.00	2.20
Interviewer Level	22.29	22.65	7.52	21.97	21.55	5.66
Regional Level			22.65			21.54
ICC	0.13	0.11	0.10	0.11	0.07	0.06
Respondents	11,058	8,154	8,154	4,864	3,821	3,821
Interviewers	307	222	222	227	163	163
Regions			6			12
Avg N/group	36.0	36.7	1359.0	21.4	23.4	318.4

Source: U.S. Census Bureau. Survey of Income and Program Participation-Event History Calendar (SIPP-EHC), 2010 and 2011.
†—Not applicable for the 2010 SIPP-EHC.
*p-value ≤ 0.05; **p-value ≤ 0.01; ***p-value ≤ 0.001

Figure 1. 2010 and 2011 SIPP-EHC Interviewer Learning Curve



Source: U.S. Census Bureau. Survey of Income and Program Participation-Event History Calendar (SIPP-EHC), 2010 and 2011.