

Certifiable?
Using Paradata to Evaluate Fields Representatives' Performance
in the Survey of Income and Program Participation

Jason M. Fields
Matthew C. Marlay
U.S. Census Bureau¹

Brief Abstract (140 Words)

The Survey of Income and Program Participation (SIPP) has begun administering a certification test to its field representatives (FRs), prior to their being allowed to conduct interviews with survey respondents. As this paper shows, certification test scores directly correlate with field performance, including respondent refusal rates, interview timing, and interview completion percent. Next, using the certification test results in conjunction with other paradata (e.g., FR experience), we examine how results from the 2013 SIPP field test are affected by variations in FR ability. Additionally, we attempt to predict how training success as measured by certification results will affect results from the first wave of the 2014 SIPP panel. Finally, we use contact history data and both consent responses and recordings generated by our Computer Assisted Recorded Interviews (CARI) to compare results from interviews with the broader predictions from the paradata.

Background

The Census Bureau has spent the past five years reengineering its Survey of Income and Program Participation (SIPP). In 2006, because of budgetary conditions that threatened the cancellation of the SIPP program, the Census Bureau, with significant support from the SIPP stakeholder community, began the re-engineering process that resulted in the instrument being fielded for production in 2014. The key motivations for the change to the existing survey design were related to both budget and respondent burden. We decided that the best way to approach this challenge was to redesign the survey from the ground up.

We will now administer the survey was through a series of annual interviews, revisiting sample households once per year instead of every four months. Moving to an annual interview allowed us to fold in the disparate parts of the old SIPP interview into a single instrument. Previously, the survey had a set of “core” questions,² which we asked during each interview (i.e., every three months). Additionally, each wave then had additional questions on specific topics, contained in a series of “topical modules.”³ For example, a Child Care topical module asked a number of

¹ This work is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views or opinions expressed in the paper are the authors’ own and do not necessarily reflect the views or opinions of the U.S. Census Bureau.

² Core questions are typically asked at the start of the interview. At the beginning of each household visit, the Census Bureau interviewer completes or updates a roster listing all household members, verifies basic demographic information about each person, and checks certain facts about the household. Core questions covering key areas of SIPP follow the initial questions and include Demographics; Labor Force; General Income; Assets; Health Insurance; Education; and Program Participation.

³ See http://www.census.gov/sipp/top_mod/topical.html for the list of SIPP topical content.

questions about the respondents' use of different forms of childcare, the costs of each form, etc. Some of these topical modules were repeated during the life of the panel, while others were only asked in a single wave. Now, with the redesigned survey, the core and topical module questions have been combined into a single interview.

We accomplished this combination via the introduction of an Event History Calendar (EHC) to the SIPP (Fields and Callegaro, 2007). This was an important addition for us, because with the larger gaps between interviews, we are now expecting respondents to recall events that may have happened at any point during the previous year, as opposed to only over the previous four months, as in the old SIPP. Additionally, the EHC design allows for a style of conversational interviewing that had previously been impossible; now, the Field Representative (FR) can discuss different domains in the respondent's life in the order that the respondent wishes to discuss them (Belli, 1998). For example, if the respondent wants to discuss his employment and then his health insurance (as might be logical, if his health insurance came from his employer), the EHC design allows this. Another advantage of the EHC design is that it allows the FR to use one area to clarify responses in another. For example, if the respondent reported losing his job in July, but cannot remember when he lost his health insurance, the FR can look back at the calendar and probe, "Was it in July, around the time your employment ended?" However, this greater flexibility also requires the FRs to be more involved in the interview process, and to retain more information about each respondent throughout the course of the interview.

Training for interviewers is critical in any survey (Groves, et. al., 2009, Chapter 9). In a survey in which more flexible interviewing methods are employed, like the EHC navigation and probing, interviewer training becomes even more important as a tool for reducing interviewer related bias and standardization of non-standard procedures. Teaching both new and experienced interviewers to interview using the conversational probing in the EHC and using standard read-as-worded methods through the majority of the detailed questions is a complex task. The two kinds of interviewers present different challenges related to adopting new methods and respecting the long-held methods to reduce interviewer effects and bias. To that end, and following the advice and guidance from the survey teams of several non-federal surveys, SIPP has begun administering a certification test to its field representatives (FRs), prior to their being allowed to conduct interviews with survey respondents.

Field Tests

In order to test the efficacy of these changes, we have conducted a series of field tests of the redesigned SIPP instrument. In 2008, we conducted a brief, paper-based field test using a subsample of respondents from the 2008 SIPP panel. In 2010, we began the first field test of the integrated SIPP instrument with an event history calendar (SIPP-EHC). Armed with promising results from both the 2008 and 2010 field tests, we moved ahead with a multi-wave dress rehearsal. This dress rehearsal began in 2011, with a Wave 1 interview, with subsequent interviews in 2012 (Wave 2) and 2013 (Wave 3). Besides evaluating the survey instrument itself, we also used the various field tests as a forum to evaluate our procedures for certifying the field representatives and collecting paradata about the cases.

Certification Test

First, we created a certification test that is administered to the field representatives at the end of their weeklong training session, conducted approximately 2-4 weeks prior to the start of the interview period. The certification test quizzes the FRs along a number of dimensions designed to test their knowledge of three areas: SIPP content, instrument navigation, and field procedures.

We test about SIPP content because we feel it is vital that the FRs be knowledgeable about the topical content on the SIPP, in case they need to respond to question from the respondents or to clarify commonly mistaken nuances in the survey questions. For example, although Social Security and Supplemental Security Income (SSI) are separate government programs, recipients often are not sure which they receive, and so report SSI as Social Security. Therefore, we include a question about the difference between the two on the certification test so we can be sure that the FRs understand the difference.

The instrument navigation questions are just that: they cover how to do common things in the instrument, such as deleting an incorrect spell in the Event History Calendar or when to provide the Flashcard Booklet to the respondent. Finally, the field procedures questions ask the FRs how to perform common tasks that they will be called upon to do in the field, such as locating and interviewing movers, classifying noninterview situations properly, and so on.

As with the SIPP-EHC field test, the certification test was originally conducted via a paper questionnaire. Some of the initial findings as related to the development of an interviewer certification test for the first field test were presented in a poster at a previous Population Association meeting (Walsh, 2011). For the 2010 SIPP field test, we found that test scores had a non-significant, but mediating, effect on interview length and a significant effect on nonresponse rates (explaining 7.5% of an interviewer's overall non-response rate). In a combined model, certification test scores could explain 13.7% of variability in interviewer productivity. Through the rest of the field tests, the interviewer certification test was refined and automated, and in the 2014 production panel, we are administering the certification test to FRs via a Blaise instrument questionnaire. This automation reduces the potential for cheating on the test (as we can now capture keystroke files, total time spent, etc.) and gives us an easily analyzable dataset (previously we had to manually code the test results into a SAS dataset, which is both time-consuming and prone to entry error).

Paradata

Second, in addition to the results from the certification test, we also collect other paradata from many sources. There are data from a variety of interview management systems that the Census Bureau uses, as well as information contained in the SIPP survey instrument itself. For the interview itself, we collect information such as the length of time it takes to complete, section lengths, don't know and refusal use, consent to record, calls for help screens, and many others. From a contact history instrument, data are collected about the interview such as where it was done (inside/outside, in which room of the house, etc.), the number of contacts required, observations about the unit and neighborhood, and respondent attitudes like reluctance or hostility. For the field representatives, we collect their experience with SIPP, experience with

other Census surveys, seniority, region, and so on, however, basic demographics of the interviewers are not available by policy.

Recording Interviews

Our final source of information about the interview comes through the Census Bureau's Computer Assisted Recorded Interviewing (CARI) software. CARI records the complete interaction between a field representative and a respondent for a set question. It also captures a screenshot of the question after the FR enters the respondent's answer into the instrument. These recordings and screenshots are stored on the FR's laptop and transmitted securely to Census headquarters at the end of each day, when the FR transmits all of his interview results. Headquarters staff can then analyze these data to see both how the FR is performing and whether there are any areas of the instrument that are causing problems (such as questions that respondents have difficulty interpreting). The CARI system does not record the entire SIPP interview, as doing so would provide us with such a massive dataset as to render it unusable. Instead, we are recording a small subset (10-12 questions) of each respondent's interview.⁴

Data and Analyses

Using all of these sources of paradata and other information about the interview, this paper presents a number of analyses. First, we show results from the certification test scores and analyze how highly they correlate with field performance, including respondent refusal rates, interview timing, and interview completion percent. Our hypothesis is that higher scores on the certification test (i.e., the more knowledgeable the FR is about SIPP and the SIPP-EHC instrument) will lead to better outcomes in the field.

Next, using the certification test results in conjunction with other paradata (e.g., FR experience), we examine how results from the 2013 SIPP field test are affected by variations in FR ability. For example, it seems intuitive to believe that FRs who have a lot of Census experience should have higher overall response rates because they have learned how to convert respondents who initially refuse to participate into completed interviews. On the other hand, because the SIPP-EHC instrument is so different from both previous SIPP instruments and from other Census surveys, it may be the case that new FRs actually have an advantage, because they have no preconceived notions about how the survey is supposed to work. The paradata will allow us to test this question.

Additionally, we attempt to predict how training success as measured by certification results will affect results from the first wave of the 2014 SIPP panel. Using the 2013 field test data, we should be able to build a model that assesses how well certification test scores affect outcomes. With this model in hand, we can apply it to the 2014 certification test results – we are administering the 2014 training in December 2013 and January 2014, so we will have a complete dataset of the certification test scores before we will have any data from the 2014 SIPP

⁴ Note that CARI works on a person level, not on a household level, so we will obtain CARI data from each respondent individually. Also, we obtain the respondent's consent before recording him or her, so some respondents may opt out of being recorded. CARI refusal rates are another piece of paradata that we can incorporate into our analyses.

interviews. So, while we will not be able to show any actual results from 2014 by the time of the PAA meetings, we should be able to show our predicted results.

Finally, we use contact history data from our Contact History Instrument (CHI) and recordings generated by our Computer Assisted Recorded Interviews (CARI) to compare results from interviews with the broader predictions from the paradata. For example, the CHI collects the number of contact attempts it takes before an interview is completed, any reluctance on the part of the respondent, the methods of contact, and observations by the FR about the unit and neighborhood. This will allow us to analyze if, for example, having to make multiple visits to a household leads to poorer quality data being collected, or if the additional visits allow respondents time to gather records and provide us with better data. The CARI recordings will serve as a useful check on assumptions about FR experience and preparedness. For example, if we encounter a question that has a high rate of refusals, we can listen to a sample of the recorded answers to that question to see what the FR/respondent interaction was like when the question was being administered. Ideally, we will be able to have some summary measure from a quality assurance review of all of the recordings, but at this point, there are resource constraints associated with listening and coding all the recorded interviews.

Conclusion

Paradata is a relatively new tool in the Census Bureau's arsenal. As SIPP is one of the first Census surveys to make extensive use of the paradata available to us, we believe that the results shown in this paper will make a real difference in the way that Census – and possibly other survey agencies – handle their survey administration, especially in an age of constrained budgets.

References

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

Fields J.M.; Callegaro M. "Background and Planning for Incorporating an Event History Calendar into the Re-Engineered SIPP." Federal Committee on Statistical Methodology, Alexandria, VA; 2007.

<http://www.census.gov/sipp/DEWS/Fields%20and%20Callegaro%20FCSM%20Final%20Paper%202007.pdf>

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. (2009). *Survey Methodology*. Hoboken, NJ: John Wiley and Sons.

Walsh, R. (2011). "Certification Tests as an Indicator of Interviewer Productivity in the 2010 SIPP-EHC." Poster presented at the Population Association of America Meetings. Washington DC.