

# **Analysis of Recorded Interviews in the 2010 SIPP-EHC Field Test: Executive Summary**

Jeffrey Moore  
Contractor/Consultant  
Social, Economic, and Housing Statistics Division

SEHSD Working Paper Number 2012-17

August 2, 2012

## Introduction

The US Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP). A key feature of the re-engineering effort is a shift from a rolling four-month reference period to annual data collection focused on the preceding calendar year, using event history calendar (EHC) interviewing methods. As part of its long-term research program to develop and evaluate the use of EHC methods, the Census Bureau carried out the 2010 SIPP-EHC Field Test, which used an electronic EHC instrument to capture data about calendar year 2009. The field test sample consisted of 7,982 addresses drawn from high poverty strata in 10 states. In order to better understand how EHC-style interviewing actually unfolds under “live” interview conditions, and the effectiveness of the field representative (FR) training program, the field test included the recording of 173 interviews, 138 of which were ultimately included in this analysis. The 52 FRs who conducted the recorded interviews were comprised, in approximately equal measure, of those with SIPP interviewing experience, with interviewing experience on other Census Bureau surveys (but not SIPP), and new hires. Tucson Telephone Center (TTC) staff transcribed the recorded interviews, which were then coded at headquarters.

## FRs’ Performance on Basic EHC Interviewing Practices

The calendar aid – a paper document – was included in the 2010 field test in order to give respondents a sense of SIPP interview content, and to serve as a potential memory anchor. FRs introduced the calendar aid to respondents reasonably well – in about three-quarters of the cases the introduction was judged to be at least acceptable, if not better. However, their follow-up to encourage respondents’ active use of the calendar as a memory aid was almost nonexistent.

Borrowing from the psychology of autobiographical memory, and following common EHC practice, the 2010 field test included an opening question about landmark events. For the most part, FRs carried out the initial phase of this part of their EHC interviewing assignment quite well – about 80 percent of their introductions of landmark events to respondents were acceptable or better. About one-third of the time respondents’ immediate response was to produce a landmark event report, as desired – an outcome that was more likely following a high quality introduction (44%), and less likely following a poor one (13%). Conversely, problematic responses (e.g., requesting clarification) were more likely following poor introductions (75%), and less likely following good introductions (30%). After the introduction, most FRs (about

60%) probed for more (or any) landmarks as the training program had instructed them. Although actively poor probing for more/any landmarks was a rare event, passively poor performance – failing to probe when the circumstances clearly called for it – happened in about one-quarter of the recorded interviews. FRs’ follow-up performance in attempting to elicit landmarks was notably poor in two areas. First, although training stressed the importance of probing specifically for early-in-the-year landmarks, FRs mostly failed to do so – about 90 percent of the time when the situation called for such probing there was none. Second – and again, counter to their training – FRs usually (about three-fourths of the time) incorrectly re-asked respondents for a new set of landmarks in proxy interview situations.

Respondents produced at least one landmark event report in about two-thirds of the recorded interviews; the maximum was six. Across all interviews, respondents reported an average of 1.3 landmarks; respondents who reported any landmarks reported about two, on average. Detailed analysis finds a clear, positive relationship between the overall quality of the FR’s presentation of the landmark task and the number of landmarks produced by the respondent ( $r=+.37$ ). About one-third of respondents’ landmark event reports involved substantive EHC “domain” events that would likely have arisen during the interview without any questioning about landmarks. Conversely, about two-thirds were unique to the landmark topic, and would not have surfaced had there been no special probing for them at the beginning of the EHC interview.

#### FR Characteristics and Performance of Basic EHC Tasks

Scores on the EHC-specific component of the end-of-training certification test were significantly predictive of FRs’ subsequent performance on the basic tasks of the SIPP-EHC interview: the calendar aid introduction, the introduction of landmarks, landmark probing, etc. There was no such relationship for the SIPP-specific sub-scale of the test. Relationships between performance of the basic EHC tasks and FR experience/tenure were not as obvious, although the results suggest a slight tendency for better performance among experienced (non-SIPP) FRs and worse performance among the new hires, with SIPP-experienced FRs falling somewhere in the middle.

#### Respondent Recall Difficulties and FRs’ Responses to Them

Overt signs from the respondent that he or she was having trouble recalling the timing of some EHC-relevant event were rare. About 85 percent of the interviews proceeded without any such signs from the respondent. Across all analyzed interviews, there were only 21 (15%) in which a respondent evidenced an underlying recall problem, and only 24 recall problem incidents in total. FRs perceived respondents’ “distress signals” only imperfectly – for only 17 of those 24 recall problem incidents was there any evidence in the FR’s behavior that he or she noticed that there might be a problem.

FRs’ certification test scores show no relationship with the likelihood that they would notice a problem. Experience/tenure “notice” differences, however, were quite large. Somewhat surprisingly, SIPP-experienced FRs performed notably poorly; they were less likely to notice respondent recall problems than were new hires, and also less likely to notice than new hires and non-SIPP experienced FRs combined.

When they noticed a recall problem, FRs almost always took some form of action to try to assist the respondent, but their actions for the most part failed to exploit the unique strengths of EHC methods. In only one instance did an FR call on an already-reported event to help the respondent recall the date of some other event. In four other cases FRs employed a weaker form of memory anchoring to assist recall, by calling the respondent's attention to standard, calendar-related phenomena, such as holidays, or seasons. The remaining two-thirds of FRs' assistance attempts were a miscellaneous mix of "other" approaches marked by the absence of useful memory cueing. The modal approach in these cases was to probe for the hard-to-recall date with a direct question on the order of: "What month was that?" Despite the very small number of cases, the likelihood that a recall difficulty would be resolved (in the form of a confident-sounding answer) was significantly related to the type of assistance the FR offered. All of the cases in which the FR addressed the respondent's recall problem with some form of memory anchoring strategy were either definitely or possibly resolved, compared to only about half of those in which the FR employed some other type of assistance.

The results offer no evidence that certification test scores were significantly related to use of the more effective, "memory anchor" assistance types. The use of memory anchors does appear to have been related to FRs' prior experience, however, with new hires performing notably more poorly than the two experienced groups. In no case did a new hire FR who noticed and acted upon a perceived respondent problem use an assistance strategy with good prospects for success; FRs with some interviewing experience used such strategies half the time.

### Tentative Conclusions

- (1) Drop the calendar aid as an active tool for respondents' use.
- (2) Landmark events are not essential to the SIPP-EHC interview. Many will duplicate material soon to surface in the main substantive domains of the EHC interview. In addition, there simply does not seem to be much call for their use. Dropping landmarks does not leave the interview empty of potential memory aids since any event recalled with confidence can serve that purpose.
- (3) If landmarks are retained in subsequent iterations of the SIPP-EHC, the results highlight the need for improved FR training in several areas, including: (a) responding humanely to negative and even tragic landmark event reports; (b) probing to elicit landmarks from the early months of the reference year; and (c) what to do about landmarks in a proxy interview.
- (4) Overt evidence of respondent recall problems in the SIPP-EHC interview appeared relatively rarely. A major worry going into the test was that the new and much longer 12-month recall period, compared to SIPP's traditional 4-month period, might be very burdensome to respondents, and might result in major decreases in data quality. On the surface, at least, in the observable behavior of the respondents, those fears were not borne out.
- (5) When evidence of respondent recall problems arose, FRs too often failed to take notice, and, when they did notice, too often failed to take effective action. Future iterations of the FR training program need to focus on improving FRs' ability to recognize recall problems, and especially on better equipping them to call on effective strategies to assist recall. These

conclusions hold for all FR experience levels, but probably especially for those with no prior interviewing experience. Improvements to the instrument to render it easier and more obvious for FRs to use may offer an important route to better training, and better performance in the field.

(6) The EHC-specific component of the end-of-training certification test was positively associated with a number of basic, desired EHC interviewing behaviors. This was not the case, however, for the more complex and demanding behaviors that form the core of successful EHC-style interviewing – noticing respondent recall problems and exploiting EHC design features to assist recall. A post-training test that was predictive of these skills would allow trainers and survey managers to focus follow-up/remedial training efforts on FRs most in need of it.

(7) In order to better assess improvements in instrument design and training – especially as they affect how interviews unfold in the field – incorporate interview recording in future tests of SIPP-EHC procedures.

# Analysis of Recorded Interviews in the 2010 SIPP-EHC Field Test<sup>1</sup>

Jeffrey Moore, Ph.D.  
Contractor/Consultant  
Social, Economic, and Housing Statistics Division

SEHSD Working Paper Number 2012-17

August 2, 2012

## Abstract

The 2010 SIPP-EHC Field Test was the Census Bureau's first large-scale test of event history calendar (EHC) interviewing procedures using an electronic instrument. Current plans call for the eventual use of these procedures, as refined through a series of field and other tests, in a planned new ("re-engineered") Survey of Income and Program Participation (SIPP) survey scheduled to go into production in 2014. In order to maximize information yield, the 2010 field test included a number of special features and activities, among which was the recording of some of the interviews. This report describes those procedures, their implementation outcomes in the 2010 field test, the methods used to analyze the recordings, and the results of those analyses. Key among the research findings are the following:

- interviewers ("field representatives" (FRs)) performed quite well at introducing the calendar aid and the "landmark events" task, although their follow-up on the former – encouraging respondents to actually *use* the calendar aid – was almost nonexistent;
- about two-thirds of all respondents produced at least one landmark event report, and FRs recorded those reports very accurately;
- FRs' performance on an end-of-training "certification test" – in particular, their performance on a sub-set of test items dealing specifically with EHC procedures – was predictive of the quality of their handling of basic EHC interviewing tasks;
- to a lesser extent, FR experience/tenure was also related to basic EHC performance;
- respondents exhibited overt evidence of recall difficulty – the signal for FRs to employ EHC interviewing methods to assist recall – only rarely;
- FRs "caught" those distress signals imperfectly; about 30% of the time when a respondent's behavior suggested a recall problem, FRs did not appear to notice;

---

<sup>1</sup> This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. All views expressed are the author's and not necessarily those of the U.S. Census Bureau. This work has benefited greatly from the contributions of many others at the Census Bureau, including especially the members of the Re-Engineered SIPP Research Group. Several SEHSD colleagues provided key technical assistance with data analysis – in particular, I thank Bri Hillmer, Lindsay Monte, Sharon O'Donnell, and Rachael Walsh for their help. I also extend thanks to those who reviewed and commented on an earlier draft of this paper, especially Jamie Choi, Jason Fields, Matthew Marlay, Ann Marie Middleton, Lindsay Monte, and Joanne Pascale. Their comments and suggestions led to many improvements; responsibility for the paper's flaws, however, rests solely with me.

- when FRs did perceive respondent recall difficulties, the actions they took to assist almost never tried to exploit the supposed strengths of EHC methods, and were of generally low quality;
- FRs' end-of-training certification test scores predicted neither whether or not they noticed respondents' recall difficulties, nor, when they did notice, the quality of their assistance efforts;
- FR experience/tenure was predictive of both noticing and taking positive action to assist respondent problems, although the nature of the former relationship, in particular, does not lend itself to easy explanation;
- the quality of FRs' assistance efforts was positively related to the likelihood of an apparently successful resolution of respondents' recall difficulties.

The results of the investigation point toward a number of recommendations concerning possible SIPP-EHC procedural modifications and aspects of FR training which need to be strengthened.

## **1. 2010 SIPP-EHC Field Test Project Overview<sup>2</sup>**

### **1.a. The re-engineered SIPP**

The US Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP) to accomplish several goals, including reducing burden on respondents, reducing program costs, improving accuracy, improving data timeliness and accessibility, and improving relevance. The main objective of the survey remains unchanged – providing accurate and comprehensive sub-annual data, from a nationally representative sample, about the income and program participation dynamics of individuals and households in the United States. Key among its specific analytic goals are to enable the evaluation of annual and sub-annual income dynamics, movements into and out of government transfer programs and their causes and consequences, the family and social contexts of individuals and households, and interactions among these items. A major use of the SIPP has been to evaluate eligibility for, and “take up” of, government programs, their effects on economic (and other) well-being, and the potential impacts of proposed modifications to them. See [www.census.gov/sipp](http://www.census.gov/sipp) for a detailed description of the current SIPP program. Details regarding the re-engineering of SIPP can be found at [www.census.gov/sipp/dews.html](http://www.census.gov/sipp/dews.html).

A key feature of the re-engineering plan is a shift from a rolling, every-four-months data collection schedule, and a rolling four-month reference period, to annual data collection focused on the preceding calendar year. Despite this major structural change, the core content of the survey will remain largely the same. To accomplish the shift to annual data collection, and to facilitate the collection of monthly data over a longer reference period, the Census Bureau proposes to employ event history calendar (EHC) interviewing methods (Fields and Callegaro, 2007).

---

<sup>2</sup> Section 1 draws heavily from Fields and Walsh (in preparation).

The shift to annual interviewing and a much longer reference period has raised concern about potential degradation in the quality of monthly estimates due to recall errors (e.g., National Research Council, 2009). Belli (1998), however, provides a strong theoretical rationale for the use of EHC methods, and their likely superiority to more traditional survey instruments – the fundamental notion being that EHC methods improve respondents’ ability to integrate memory across topic areas, and to retrieve related information in a more natural, autobiographical manner. Belli and his colleagues (e.g., Belli, Shay, and Stafford, 2001) have reported research results which generally support the theory. But there are exceptions to those positive results, and the research base of strong quantitative evaluations of theory-based predictions of EHC behavior remains somewhat limited. In addition, no research has addressed the key practical issue for the Census Bureau and for SIPP’s users – namely, whether a 12-month EHC approach can yield the same data quality as a series of standard, 4-month reference period questionnaires<sup>3</sup>. Thus, concern lingers about the data quality implications of the re-engineered survey.

### 1.b. The 2010 SIPP-EHC Field Test

The Census Bureau initiated a long-term research program to develop and evaluate the use of EHC methods in the re-engineered SIPP with the 2008 SIPP-EHC Paper Test, which was primarily a proof-of-concept exercise. This test administered a 12-month, calendar year 2007, paper-and-pencil EHC instrument to a small sample of expired 2004 panel SIPP sample addresses in two states. For continuing residents at those addresses – those who had not moved since their final 2004 panel interview – we compared the SIPP-EHC Paper Test data with the data they had already reported about calendar year 2007 via the final three waves of the 2004 SIPP panel. The results from this initial investigation were sufficiently positive (Moore et al., 2009) to provide a “green light” for additional work, specifically in the form of the development of an automated EHC instrument for the 2010 SIPP-EHC Field Test which is the subject of this report.

The 2010 SIPP-EHC Field Test employed the first electronic EHC instrument created by the Census Bureau. The field test sample consisted of 7,982 addresses drawn from high poverty strata in 10 states. FRs who conducted the interviews were affiliated with six of the Census Bureau’s 12 regional offices – Boston, Chicago, Dallas, Los Angeles, New York City, and Philadelphia. Following a week-long training session (which took place in December 2009 for most FRs), the field period for the 2010 test was January through April of 2010. Of the 7,982 original sample addresses, 1,626 were found to be ineligible. The remaining 6,356 eligible addresses yielded 5,207 completed interviews, for a response rate of approximately 82 percent. SIPP’s procedures call for the completion of individual interviews (by self-response if possible, by proxy if necessary) with all adult (age 15+) members of sample households. Field test interviews enumerated 11,053 adults in the 5,207 interviewed households, 89 percent of whom provided a complete or partially complete individual interview.

---

<sup>3</sup> Most applications of EHC methods involve very long recall periods – several years, decades, even entire lifetimes – much longer than the calendar year reference period of the re-engineered SIPP interview. Somewhat ironically, therefore, another question yet to be addressed by research is whether EHC methods can be effective with a reference period as *short* as a year.

## 1.c. Overview of this report

Census Bureau staff have carried out a number of evaluations of the quality of the data from the 2010 field test – see, for example, Edwards (2012), Ellis and Chenevert (2012), Marlay and Mateyka (2012), SIPP-EHC Data Evaluation Workgroup (2011), and Stinson, et al. (2012). The focus of the current investigation is very different. It seeks to understand the quality of FRs’ implementation of key features of the new (for SIPP) EHC interviewing methods, through an evaluation of 2010 field test interviews recorded for that purpose. The remainder of the report is organized as follows: Section 2 describes the recording and transcription procedures. Section 3 briefly summarizes the coding of the transcripts, and the selection of the final set of transcripts for analysis. Section 4 examines the prior interviewing experience and end-of-training certification test performance of field test FRs, and compares those who recorded interviews and those who did not record. Section 5, the first substantive results section, uses the transcripts to assess FRs’ implementation of basic EHC interviewing practices; section 6 assesses the relationship between those EHC performance measures and FRs’ experience/tenure and certification test scores. Section 7 uses the recordings to identify instances of respondent difficulty in recalling the dates of events of interest to the survey, and to examine FRs’ responses to those difficulties. Section 8 offers a set of tentative conclusions, based on the research findings, concerning modifications to EHC interviewing procedures in future iterations of the SIPP-EHC questionnaire, and areas where enhancements to FR training are needed<sup>4</sup>.

## 2. Field Test Interview Recording and Transcription

The Re-Engineered SIPP Research Group, convened by the SIPP Executive Committee, and comprised of staff representing all of the Census Bureau’s SIPP-participating divisions, was charged with designing the 2010 field test. The Research Group recommended that interview recording be incorporated into the test design in order to address several research goals, the most important of which were intended to yield a better understanding of how key features of EHC-style interviewing actually unfold under “live” interview conditions. Did FRs implement the “calendar aid” procedures as intended? How well did they introduce and probe for landmark events? How did respondents respond to the landmarks reporting task? Was there any resistance or reluctance on respondents’ part to report landmarks? Did FRs record respondents’ landmarks accurately? How much “demand” was there for the use of landmarks and other information to assist the response process – that is, how often did respondents signal that they were having difficulty recalling the date of some relevant event? When those signals arose, did FRs notice them, did they take action, and, if so, was it proper action? And so on. The underlying goals were to assess the EHC procedures themselves, as well as (by deduction) the effectiveness of the training program administered to FRs to ensure that they were reasonably adept at implementing those procedures.

---

<sup>4</sup> As suggested earlier, this analysis only examines the Census Bureau’s inaugural launch of an automated EHC instrument and its associated training. Refinements have been implemented to both the instrument and the training for subsequent rounds of field testing – in particular, a revised wave 1 test conducted in early 2011, and a wave 2 test among those same households, and incorporating dependent interviewing procedures, fielded in late spring/early summer of 2012.

## 2.a. Confidentiality concerns

Although interview recording is a common procedure in the survey world in general, outside of relatively small-scale, laboratory-like research settings there is very little precedent for it at the Census Bureau. Thus, the prospect of recording interviews in the field aroused some of the wariness that can attend almost any new endeavor. But quite specific concerns about confidentiality also surfaced in the earliest discussions about recording – very understandable concerns, it should be noted, with which the research planners themselves were quite sympathetic. Data confidentiality is serious business at the Census Bureau. Confidentiality is, first of all, a simple and direct promise made to those who provide the data, and is important for that reason alone. But it is also an axiom of Census Bureau culture that the agency's very ability to operate in the public realm rests on a foundation of voluntary cooperation, and that a confidentiality failure could damage or even destroy that ability. The recordings aroused even greater-than-normal sensitivity and concern about disclosure risk. Some of that additional concern was due to the nature of the data – not just sterile computer keystrokes, but respondents' (and FRs') actual words and voices. Also contributing to a heightened concern was the timing of the field test, on the eve of the 2010 Decennial Census, when even a hint of a confidentiality breach could potentially be catastrophic. For all these reasons, the Research Group worked carefully and thoroughly, for many months, to inform and seek feedback about plans to record interviews in the 2010 field test from all of the many interested areas of the Census Bureau, including the legal, IT security, and policy offices, and even the agency's executive committee.

## 2.b. Recording plans and procedures

One key factor that made recording possible for the 2010 field test was the serendipitous timing, a few months before the test, of the wholesale replacement of the entire existing stock of rather outmoded computer-assisted interviewing equipment with a new generation of laptop computers. Most importantly for present purposes, the new devices came with a built-in digital recording capability. The old laptops lacked this capability, which meant that any recording effort would have required that FRs carry around and manipulate a second piece of machinery, in addition to the laptop itself – a huge and very intrusive additional burden for field staff. Not only did the recording feature of the new laptops overcome those physical problems, it also solved many data security issues. Because the recorders were incorporated into the laptops themselves, the sound files they produced were automatically protected by the same strong data encryption and password protection safety measures used to ensure the confidentiality of the interview data themselves. Finally, an additional benefit of the new laptops was a greatly increased memory/storage capacity, which technology staff determined was more than sufficient to hold a far greater amount of sound file data than the maximum load that plans called for [see below], and to do so without any noticeable drag on the operating speed of the automated questionnaire.

The wide range of research issues that the Research Group hoped to address with the recordings led the group to recommend the recording of entire interviews, rather than brief, selected sections. By the time recording plans were formulated it was deemed to be too late in the instrument development process to include automatic operation of the recorder in the questionnaire, and this also fed into the decision to record everything. The on-off operation of the recorder had to be a manual process, and recording the whole interview simplified the actions

required by FRs. Staff in the Technologies Management Office (TMO) developed procedures whereby, with a few simple keystrokes, FRs could start and end the recording. Included in those procedures were automated processes that assigned a unique name to each sound file, incorporating as part of each file name the date and time of the start of the recording, the case id, and an identifying code for the FR conducting the interview.

In addition to addressing basic questions about the administration of the new EHC procedures, the Research Group also hoped to be able to use the recordings to document change across time in how FRs administered those procedures. Specifically, we expected that FRs' performance might improve over the course of the field period as they became more familiar and comfortable with all aspects of the interview<sup>5</sup>. We therefore implemented a system of three recording "windows" – one at the beginning of the field period, another about a month in, and a third at what we guessed would be near the end of interviewing. Because we did not sample cases to be interviewed at each of these times we knew that selection biases would cloud any analysis of time trends, but we hoped to be able to draw some tentative conclusions, at least, about performance changes over time. The Research Group planned the recording window dates in advance, but did not reveal them in advance to FRs. At the opening of each window, an announcement went out from each RO to each participating FR [see below] that said, in effect, "Starting now, record the next two interviews you conduct." Upon receiving that message, FRs knew that the next time they were in a household and ready to start an interview with a cooperative respondent, they should record that case and the next one like it. Thus, we expected that each participating FR would produce six recorded interviews – two at the beginning of the field period, two in the middle, and two near the end.

The first task in recording an interview was to gain the respondent's consent; no recording was permitted without that consent. To confirm the consent, it was recorded; FRs asked the respondent whether they could record the interview, obtained consent, then turned on the recorder and re-asked the consent question, in order to capture the respondent's consent on the recording. SIPP procedures generally treat 15-year-olds as adults. However, to avoid any issue having to do with a person's age-related ability to give consent, we implemented an age restriction on recording such that only persons aged 21 or older were eligible to be recorded. Many of the other specific recording procedures were designed to keep the "rules" as simple as possible for FRs. For example, we made no attempt to sample respondents for recording in to-be-recorded households, but instead instructed FRs simply to record the interview of the first person to be interviewed for as long as that person continued to provide interview data. That is, recording was to capture the portion of the interview dealing with the household as a whole, the person's own, individual, self-response interview, and any additional interviews that same person provided as a proxy respondent for other members of the household. For the sake of simplicity we wanted to avoid asking consent multiple times in a household, so only one respondent was to be recorded; as soon as a second person entered the scene to provide his or her own self-response interview, recording was to stop. Also for simplicity, we established the rule that recording was not to be turned back on in a household once it had been turned off. Thus, if an interview that was being recorded was interrupted, requiring a return visit to complete the interview, the remaining portion of the interview was not to be recorded.

---

<sup>5</sup> Walsh and Fields (2012) find evidence consistent with this notion in their evaluation of the association between interview length and interview order.

The final major hurdle that the recording plan presented was how to get the recorded interview data files from the laptops to secure storage. FRs use an electronic system to transmit standard interview data from their laptops to secure data storage in each of the Census Bureau's regional offices (RO). The size of the sound files caused some concern among technology staff about using the same system to transmit both types of files. There was also concern about the potential vulnerability of recordings that would spend a lot of time "walking around" out in the world inside laptops that do occasionally get stolen or otherwise go missing – data encryption and password protection notwithstanding. So the goal was to remove the sound files from the laptops sooner rather than later. We devised a relatively "low tech" solution to these problems. We required that, at some reasonably convenient time within a few days of recording an interview, FRs bring their laptops to the RO and have the recording data file stripped off the laptop and directly into central storage via a secure cable connection.

This plan, and in particular the cost implications of the additional travel it required, led to a geographical restriction on FR eligibility to record interviews. At the suggestion of Field Division (FLD) staff, we adopted a "50-mile" rule: eligible-to-record FRs must live or work within 50 miles of their RO, since by RO rules travel longer than 50 miles requires paying "per diem" costs in addition to mileage and work hours. (One effect of this rule was to eliminate the Philadelphia RO from participating in the recording, since all of the cases included in the field test from within the Philadelphia RO's boundaries were located in Maryland.) FLD's initial estimate was that 136 of the FRs who were going to be involved in the field test would meet that criterion. With the plan of six recorded interviews per participating FR, this level of FR participation would have produced over 800 recordings, far more than the 500 that the Research Group had roughly targeted as a feasible volume to handle and analyze. Thus, we requested that FLD reduce FR participation by selecting only about 60% of the total number of FRs eligible under the 50-mile rule, in order to yield approximately 80 to 85 recording FRs, and FLD agreed to do so. (The Research Group suggested that the ROs select the FRs closest to the RO, but made no attempt to assess how, in fact, RO staff made their selections.) FLD's final estimate of FRs designated as eligible-to-record was 82; under the plan of two recorded interviews per FR in each of the three recording windows, that staffing level would yield approximately  $(82 \times 2 \times 3 =)$  492 recorded interviews.

### 2.c. Recording outcomes

As shown in Table 2.c.1 (next page), the actual production of recorded interviews fell far short of the Research Group's plan. Although most of the ROs met or came close to their projected overall level of FR participation – NY being the notable exception – in the end the number of FRs who recorded any interviews (56) was only about two-thirds of what was expected, and those FRs who did record only recorded an average of a little more than three interviews apiece, about half of the expected output of six. (In fact, only three FRs produced the expected output of six recordings.) The combined result of these factors was a yield of recordings (173) that was only about one-third of the projected total. In addition, as the right-hand side of the table shows, recording was sporadic across the three recording windows. Only about one-sixth (9) of the FRs who recorded any interviews recorded at least one interview in each recording window; in fact, only one FR recorded interviews exactly as planned – two recordings in each of the three windows.

---

**Table 2.c.1: Interview Recording – Planned Design vs. Actual Results**

RO	Planned		Actual		# of FRs who Recorded in...		
	# FRs	# recs	# FRs	# recs	...1 window	...2 windows	...all 3 windows
BOS	11	66	11	35	4	4	3
CHI	10	60	9	28	2	6	1
DAL	7	42	7	32	1	2	4
LA	18	108	14	38	5	8	1
NY	36	216	15	40	7	8	0
TOTAL:	82	492	56	173	19	28	9

---

Source (for this and all subsequent tables): US Census Bureau, 2010 SIPP-EHC Field Test internal files.

In retrospect, we should have anticipated the inevitable fall-off in FR participation, and in the number of recorded interviews, especially in recording window 3, by which point a large number of the total caseload should have been resolved. But FR participation and recording production was well below expectations right from the start. The detailed results (not shown) reveal that 44 FRs produced 76 recordings in window 1, as opposed to the 82 FRs and 164 recordings expected. In window 2 those numbers fell to 37 FRs and 64 recordings; and then only 21 FRs produced only 33 recordings in window 3. Even discounting window 3, however, it is still the case that only half (28) of the participating FRs recorded interviews in both window 1 and window 2 (data not shown). The small number of FRs who participated in the recording relative to what was expected, the small number of recordings they produced, and the lack of consistent and robust recording across the field period – combined with a weak research design to begin with – led to the abandonment of any attempt to use the recording data to assess changes in FRs’ performance with the new instrument and procedures over time. Instead, all analyses that follow in this report simply combine all of the recordings into a single group and ignore the recording window feature of the study.

#### 2.d. Transcription

We arranged with staff of the Tucson Telephone Center (TTC) to transcribe the recorded interviews. Research Group staff developed a 3-day transcription training package, and administered the training in Tucson in January 2010, at the same time that the first recordings began to arrive from the field. Transcriptions of the recorded interviews were to be verbatim. That is, they were to include every utterance recorded in the interview in exactly the way that the FR or the respondent produced it, regardless of bad grammar, mispronunciation, breaks/interruptions/false starts, irrelevance to the interview, absence of logic, and even the absence of words (e.g., the transcriptions were to include vocal fillers such as “mm-hmm,” “uhh,” “umm,” etc.). All was to be captured, in other words, and in as close a manner as possible to how it actually unfolded during the interview (see Pascale, Moore, and Chan, 2010).

Training focused heavily on transcription practice and review. The first day used a specially prepared mock interview with an existing “gold standard” transcription so that trainees’ initial efforts could be assessed and discussed, and re-training or clarification needs could be identified and resolved. On day two of the training, the trainer and trainees all transcribed the same portion of a “live” interview; again, the transcription practice was followed by a group discussion and review to identify and repair weak spots in the training. On the third day of training, the trainees began individually transcribing their first live cases, which they brought to the trainer for a one-on-one review. In addition to individualized instruction, this final review process also identified a few remaining general issues which were discussed with the group as a whole.

Transcription plans exploited the fact that many of the TTC transcribers were bilingual English-Spanish speakers. Thus, we applied the same rules to recorded interviews that had been conducted completely or partially in Spanish as to any other interview – the participants’ utterances were to be transcribed verbatim, exactly as they had been spoken. Some of those interviews involved a translator serving as the link between an English-speaking FR and a Spanish-speaking respondent – the FR read the questions in English, the translator translated the question into Spanish for the respondent, the respondent responded in Spanish, and the translator translated the respondent’s response into English for the FR. In other cases, where the FR was bilingual, the entire interview was conducted in Spanish<sup>6</sup>. In these latter cases, we requested from TTC staff a translation of the Spanish into English, in addition to the transcription of the interview interaction as it transpired, in Spanish. (For reasons that are not clear, this translation component of the transcription procedures did not happen.)

One final feature of the design of the transcription process deserves mention: by design, each recording was to undergo two “passes” by two different transcribers. The intent was that the first transcriber’s pass through the recording would capture the basic details of the interaction, and the second transcriber’s pass, using the first pass as a base, would implement whatever repairs and refinements were necessary to render the final product as complete and accurate as possible. As shown in Table 2.d.1, TTC staff followed this component of the process well, with about 90 percent of the recordings receiving at least two transcription passes.

---

**Table 2.d.1: Recorded Interviews – Number of Transcription “Passes”**

	<u>N</u>	<u>%</u>
1 pass only	17	10
2 passes	140	81
<u>3 passes</u>	<u>16</u>	<u>9</u>
TOTAL:	173	100

---

<sup>6</sup> The 2010 field test instrument was English-only; FRs who conducted the interview in Spanish had to construct Spanish-language versions of the questions “on the fly.” Note also that Spanish was the only non-English language to appear on any of the recorded interviews.

### 3. Coding the Transcripts

#### 3.a. The coding scheme

As noted above, the primary motivation for recording interviews in the 2010 field test was to shed light on how key features of EHC-style interviewing actually unfold under “live” interview conditions. The coding scheme I devised for these purposes was not a traditional behavior coding approach, which tends to focus on how closely interviewers follow a questionnaire script, and on the adequacy of respondents’ responses with regard to meeting question objectives. Nor was it a detailed, linguistic-type analysis of the interactional aspects of an interviewer-administered survey. Instead, the coding scheme focused on a quite specific set of questions concerning quite specific EHC procedures, including the following:

- the quality of FRs’ introduction of the calendar aid<sup>7</sup>;
- whether the calendar aid was ever mentioned again, after its introduction;
- the quality of FRs’ introduction of “landmark event” reporting;
- respondents’ initial response to the landmark events task;
- if respondents requested clarification (concerning landmarks), the quality of FRs’ response to the clarification request;
- whether FRs probed appropriately for more (or any) landmarks;
- whether FRs probed appropriately for early-in-the-year landmarks;
- whether FRs probed for landmarks in any inappropriate ways, and, if so, how often, and in what manner;
- the number of landmarks that respondents reported, and their content/subject-matter;
- the concordance between respondents’ landmark events reports (both number and content) and what the FR recorded in the questionnaire;
- the positive/neutral/negative character of reported landmarks;
- the timing of reported landmarks (month) during the reference year;
- whether, during the administration of the EHC, respondents exhibited signs of having difficulty recalling the date of some event or circumstance of interest;
- whether, when those signs of recall difficulty appeared, FRs gave evidence that they noticed the problem;
- the kinds of assistance FRs offered when they noticed a possible recall problem, and the quality of those assistance attempts; and
- whether the recall difficulty seemed to be successfully resolved.

For the most part, coding the recorded interviews on these dimensions called for subjective judgments of the nature of “chunks” of talk. For example, regarding the quality of the FR’s presentation of the landmark events task to respondents, the specific question I posed in my coding scheme was: Did the FR present the landmark task in a clear, coherent manner? The coding scheme, in essence, defined the endpoints of a three-point scale – (code 1) yes, the presentation was clear and coherent; or (code 3) no, the presentation was actively bad, neither clear nor coherent – and used a middle category (code 2, labeled here “ok/good”) for presentations that were neither actively bad nor particularly good. “Omitted” was a separate

---

<sup>7</sup> See section 5.a, below.

category. There were also non-substantive, missing-data-type codes, in case the FR's words were inaudible or had taken place "off camera," before the recording started.

Especially given the subjective nature of much of the coding, the process would have benefited greatly had there been multiple coders, operating independently. This would have enabled a statistical evaluation of the quality (reliability) of the coding, and would have permitted some refinement of the assigned codes in cases of coder disagreement. Unfortunately, I did not have that luxury. I took great pains with the coding, often returning to the sound files themselves, for example, to try to clarify difficult-to-resolve circumstances, and I tried to be very consistent in the application of rules and standards. In addition, in the results that follow, I often present verbatim segments of interview interactions that exemplify particular coding choices. I selected these segments using a random number generator to identify a particular case from among those assigned the relevant code. It is perhaps worth noting that the vast majority of these randomly selected examples stood up to scrutiny, in my judgment, in a second viewing<sup>8</sup>. In the end, however, all of the coding judgments on which the analyses to follow are based were mine and mine alone, and, notwithstanding my confidence that I did a reasonable job, this is clearly a limitation of the research.

### 3.b. Recorded interviews – final analysis subset

Of the 173 recorded and transcribed interviews, I used only about 80% in the analyses that produced the results to follow. My limited ability to understand Spanish rendered the 13 recorded interviews conducted completely in Spanish inaccessible to this project, and thus I excluded them. I also decided to exclude 13 other interviews that had used a bilingual interpreter to mediate between an English-speaking FR and a Spanish-speaking respondent. Both of these exclusions are unfortunate; there are many reasons why non-English interviews of both types are of real, substantive interest to an investigation such as this one. However, gaining insights into EHC-relevant processes in such interviews, and teasing those out from issues that arise because of the language barrier posed by the English-only questionnaire, would require an additional set of skills not available to the project. In addition to the language-related cases, I also excluded from the analysis a handful of cases for which the recording did not include the EHC portion of the interview for one reason or another. Table 3.b.1 (next page) summarizes the winnowing-down process (described above) which yielded the final analysis subset of 138 cases. For four FRs, this process removed all of their recorded interviews from the analysis subset. Thus, reducing the number of recorded interviews for data analysis carried with it a reduction in the number of recording FRs for analysis as well, from 56 to 52.

## 4. Results I: FR Characteristics – Certification Test Score and Experience/Tenure

This initial results section looks briefly at the characteristics of FRs who participated in the 2010 SIPP-EHC Field Test. I first use these characteristics to compare "recorders" with their fellow FRs who did not record any interviews. The intent is to provide some evidence to assess whether the FRs whose behaviors can be observed in the recordings were in some sense representative of

---

<sup>8</sup> Obviously, "...vast majority" is not the same as "all." In a few instances, I decided that I had coded the randomly selected case incorrectly; when that happened I repaired the incorrect code and selected another example.

---

**Table 3.b.1: Recorded Interviews – Final Analysis Subset**

<u>Total, Recorded and Transcribed Interviews:</u>	<u>173</u>
Not Usable – interview in Spanish	13
– partial interview (incomplete EHC)	2
– <u>truncated recording (incomplete EHC)</u>	<u>7</u>
Total, Potentially Usable Transcripts:	151
<u>Not Used – interview conducted with Spanish translator</u>	<u>13</u>
<b>Final Analysis Subset</b> (usable, English-language interviews):	138

---

the entire set of FRs who conducted field test interviews. To the extent that they were not representative – for example, if the processes which produced the FRs who recorded were biased toward highly capable FRs – then confidence is reduced that the findings regarding EHC-related performance that are based on the recordings are generalizable to FRs as a whole<sup>9</sup>. The variables with which to address this question are very limited. In fact, there are just two – FRs’ prior interviewing experience and their performance on an end-of-training “certification test.”

#### 4.a. Interviewing experience among recording and non-recording FRs

In discussions leading up to the 2010 field test, FLD had estimated that the staff available to conduct the test would consist in roughly equal measure of FRs with SIPP experience, FRs with decennial census or other Census Bureau current survey interviewing experience (but not SIPP experience), and new hires. The actual breakdown for all FRs was reasonably close to that 1/3-1/3-1/3 “guesstimate,” with an observed tendency (observation only; no statistical testing) for a somewhat greater representation of SIPP-experienced FRs, and a correspondingly smaller representation of experienced (non-SIPP) FRs – see Table 4.a.1 (next page), first column.

The middle column of Table 4.a.1 shows the distribution of prior experience among the 52 FRs who recorded the 138 interviews to be used in the main substantive analyses to follow; the right-most column shows that distribution for the remaining FRs (the vast majority of whom did not record). Despite appearances, a chi-square test comparing these distributions finds that they do not differ significantly (chi-square=3.57, 2 df, n.s.). Individual comparisons, however, reveal a significantly higher proportion of non-SIPP experienced FRs among those who recorded compared to the proportion with that experience/tenure level among those who did not record; other differences are not significant. Thus, the bottom line here – whether the mix of prior experience among recording FRs differed from that of FRs who did not record – remains somewhat murky.

---

<sup>9</sup> Another potential source of bias in trying to generalize from this study, of course, is that FRs who recorded interviews might have modified their behavior in those interviews because they knew it was going to be observed. The impacts of any such effects are unknown and probably mostly unknowable (see, however, section 5.c).

**Table 4.a.1: Prior Interviewing Experience –  
All FRs, and FRs Who Did and Did Not Record**

	TOTAL, ALL FRs (n=303)	FRs Who Recorded (n=52)	FRs Who did NOT Record (n=251)
<u>Total:</u>	<u>100%</u>	<u>100%</u>	<u>100%</u>
SIPP-experienced	43%	38%	44%
Other experienced (non-SIPP)	24%	35% <sup>a</sup>	22% <sup>a</sup>
New hires	32%	27%	33%

Statistical analysis summary (significant differences):

<sup>a</sup>t=1.88, 301 df, p<.10

4.b. End-of-training “certification test” scores among recording and non-recording FRs

The certification test consisted of 42 items organized into four parts: “SIPP Concepts” (which also included general survey concepts), “Landmark Scenario Questions,” “Landmark Event Procedure Questions,” and “Event History Calendar Scenarios” – see Attachment A. For analysis purposes, I collapsed the last three parts of the test into a single scale, and calculated test scores for two sub-scales, “SIPP” (25 items) and “EHC” (17 items)<sup>10</sup>, as well as a total score. Table 4.b.1 (next page) summarizes the certification test results for all FRs combined, and for the subgroups who did and did not record any field test interviews. (Note that test scores are missing for about one-quarter of the non-recording FR group and one-sixth of the recording FRs.) As the table shows, test performance means for FRs who recorded are statistically indistinguishable from those of FRs who did not record – this is true for both sub-scales as well as for the total.

4.c. Certification test scores and FR experience/tenure (all FRs)

Table 4.c.1 (next page) summarizes the certification test score results for all FRs who participated in the 2010 field test, by test sub-scale and for the total test, split out by the three experience/tenure groups. (Non-missing data for both characteristics are available for only 221 (73%) of the 303 FRs who participated in the 2010 field test.) As might have been predicted, the rank ordering of the observed mean scores on the SIPP sub-scale of the test was SIPP-experienced FRs first (20.3 items correct), and new hires last (18.3), with non-SIPP experienced FRs in the middle (19.9). (As shown in the table, statistical testing finds no difference between the mean scores for the two experienced groups; each was significantly higher than the mean for the new hires.)

<sup>10</sup> The sub-scales comprised the following items: SIPP – 1-9, 11a-11g, 14F1-14F10, 15a-15e; EHC – 10, 12, 13, 16-29. Note that EHC sub-scale items 10, 12, and 13 clearly concern EHC-specific issues even though they appear in Part 1 of the test booklet, labeled “SIPP Concepts.” In forming the sub-scales I ignored the test booklet labels and grouped items according to whether their content was primarily SIPP-focused or EHC-focused. See Attachment A.

**Table 4.b.1: Certification Test Scores – All FRs, and FRs Who Did and Did Not Record**

		TOTAL, ALL FRs (n=223)	FRs Who Recorded (n=43)	FRs Who did NOT Record (n=180)
SIPP sub-scale (25 items)	mean	19.6	19.7	19.5
	median	19.5	19.8	19.4
EHC sub-scale (17 items)	mean	12.5	12.9	12.3
	median	12.3	12.8	12.2
Cert. test TOTAL (42 items)	mean	32.0	32.7	31.9
	median	31.9	33.1	31.7

Statistical analysis summary (significant mean differences):  
[none]

**Table 4.c.1: Relationship of Certification Test Scores and FR Experience/Tenure (All FRs)**

Mean cert test score for ...	SIPP-	Other	
	Experienced FRs (n = 98)	Experienced (non-SIPP) FRs (n = 54)	New Hire FRs (n = 69)
...SIPP sub-scale (25 items)	20.3 <sup>a</sup>	19.9 <sup>b</sup>	18.3 <sup>a, b</sup>
...EHC sub-scale (17 items)	12.6	12.8	12.0
...Cert test TOTAL (42 items)	32.9 <sup>c</sup>	32.6 <sup>d</sup>	30.3 <sup>c, d</sup>

Statistical analysis summary (significant differences):

<sup>a</sup> t=4.23, 165 df, p<.0001      <sup>c</sup> t=3.48, 165 df, p<.001  
<sup>b</sup> t=2.96, 121 df, p<.005      <sup>d</sup> t=2.78, 121 df, p<.01

It is perhaps somewhat surprising that the SIPP-experienced group did not significantly out-perform the experienced (non-SIPP) FRs on the SIPP sub-scale of the test. This may in part reflect the fact that some of the “SIPP Concepts” test items actually dealt with concepts that are applicable across many Census Bureau survey activities – for example, the legal basis of the Census Bureau’s confidentiality requirements, the best response to a “why was I selected for this survey” query from a respondent, non-interview definitions, etc. Also predictable, perhaps, is the fact that the mean scores on the EHC sub-scale – which covered new material for all FRs,

regardless of experience/tenure – clustered more closely together than the means for the SIPP items, with no significant differences for any of the experience/tenure subgroup comparisons. These basic findings offer some support for the notion that the labels for the two sub-scales of the certification test are appropriate – i.e., that they measured different knowledge sets, and that each knowledge set tapped into its intended area of knowledge.

Another interesting way to look at these results is to compare performance on the two sub-sections of the certification test within experience/tenure groups. The results for the new hires suggest that under “baseline” conditions – that is, in the absence of prior knowledge – the SIPP and EHC items were similarly challenging. New hire FRs answered about 73% of the SIPP sub-scale items correctly, on average (18.3/25), and about 71% (12.0/17) of the EHC items ( $t=1.07$ , 68 df, n.s.). The prior knowledge of the two experienced groups, however, appears to have made the items in the SIPP sub-section easier to answer than the EHC items – 81% correct, on average, for the SIPP items compared to 75% for the EHC items ( $t=4.54$ , 151 df,  $p<.0001$ ). (Note that the “predictable” patterns described above are not so apparent in the results for the subset of FRs who recorded any interviews – data not shown.)

## **5. Results II: Using the Recordings to Assess EHC Interviewing Practices**

This section presents the first set of results from the analysis of the recorded interviews, the primary purpose of which was to shed light on the “black box” of EHC interviewing practices. Despite the by-now relatively common use of EHC-style interviewing, research is quite limited on how interviewers actually carry out key features of EHC methods in the field (see, e.g., Belli and Stafford, 2008). In addition, we also hoped that the recordings would help us assess the effectiveness of the EHC training program administered to FRs – the first of its kind developed at the Census Bureau – and highlight areas of that training in need of strengthening<sup>11</sup>.

### **5.a. Introducing and encouraging use of the calendar aid**

The calendar aid was a paper document whose intent was to give respondents visual cues concerning the time period and topic areas covered by the SIPP interview (see Attachment B). More importantly, it had spaces for respondents to make notes about survey-relevant events recalled with confidence, and FRs were supposed to encourage this behavior. With those notations, the aid had potential use as a memory anchor for subsequent events, addressed later in the interview, whose timing might be difficult to recall. During training, FRs received guidance about the rationale for the calendar aid and when and how to introduce it. The training message was quite simple and straightforward (see U.S. Census Bureau (2009), pages 5-3 and 5-4):

#### **When To Give The Respondent The Calendar Aid**

**Give the respondent the Calendar Aid, and a pencil, at the beginning of the calendar**

---

<sup>11</sup> As a byproduct of the transcript review, I produced informal reports on question wording/design issues and FR training issues. The former (Moore, 2011a) focuses on specific problematic questions in the SIPP-EHC instrument, excluding those related to EHC procedures. The latter (Moore, 2011b) focuses on FR training needs, with regard to both general interviewing practices and practices specific to the SIPP-EHC interview (but again excluding core EHC-style interviewing practices).

section of his or her interview. You can simply say:

FR: Here's something you might want to use to keep track of things as we go along.

#### When To Encourage Use Of The Calendar Aid

If the respondent produces some good landmark events, without too much prodding from you, start there:

R: Well, my sister got married in June – that was big. And fun. Everybody thinks she got herself a good one this time – her, too!

FR: Good – that's a good one. See that "Life Events" line? Why don't you write "sister married" in June? Anything else?

The instrument itself reinforced this guidance on a special screen, inserted just prior to launching the EHC portion of an individual interview, which read as follows:

- Distribute the Calendar Aid and pencil to the respondent so that they can follow along with the EHC interview.
- Read to the respondent as you distribute the Calendar Aid and pencil.

**Here's something you might want to use to keep track of things as we move through the interview. Please go ahead and jot down notes on this calendar as we go along.**

In coding the interview transcripts, the question I asked concerning FRs' introduction of the calendar aid was: Did the FR present the calendar aid in a clear, coherent manner? Introductions judged to be "clear and coherent" (code 1) were those which, following the suggested script, used simple and direct language to both mention the purpose of the calendar and suggest how it was to be used. At the other extreme, neither clear nor coherent introductions (code 3) included neither purpose nor use, or else used language unlikely to convey the information successfully to respondents. I applied the middle "ok/good" code (code 2) to introductions which fell short of the highest standard in one way or another, but which still seemed likely to have communicated some useful information. Table 5.a.1 (next page) summarizes the coding results, and presents examples of introductions that received each of the three main, substantive codes.

Table 5.a.1 suggests that FRs introduced the calendar aid to respondents reasonably well – in about three-quarters of the cases, I judged the introduction to be "ok/good" or better. On the other hand, the path to a "clear and coherent" introduction was both very simple and very clearly specified in the instrument, raising the question of why *all* FRs didn't avail themselves of it. In that light, omitting mention of the calendar aid entirely (presumably, in the interests of saving time), as occurred in 11 cases, makes more sense than offering a poor introduction. In four other cases, even though the introduction was not on the recording I applied a "missing" code, rather than omitted. These cases offered no evidence of the quality of the FR's introduction of the calendar aid, but did include clear evidence that it had happened, "off camera," before the recording started. As an example, in one of the cases assigned a "missing" code, the FR's first recorded mention of the calendar aid is as follows: "And you have the calendar, so we can go ahead and proceed. And you know what to do with that."

As noted above, the intent was for respondents to make active use of the calendar aid by writing down easily remembered events, which could then be referred to later, if necessary, when trying

---

**Table 5.a.1: Did the FR Present the Calendar Aid in a Clear, Coherent Manner?**

	<u>N</u>	<u>%</u>
1 – yes, clear and coherent	52	38
2 – ok/good (not actively bad)	54	39
3 – no, not clear or coherent	17	12
7 – no, FR omitted calendar	11	8
(missing)	4	3
TOTAL:	138	100%

Coding Examples (\*):

1 – FR: Ok, so there’s – I’m gonna give you that calendar aid and a pen. You might want to use it, uh, to jot down as we go along just to remember. Just to keep track...

2 – FR: Ok, I am going to give you a calendar. [pause] Now this is to help you remember important events such as change in residence, work, or other type of government programs you might participate in.

3 – FR: Ok, ahhh, let me get the calendar for you. It’s a calendar they want want you to use.

(\*) NOTE: Here and in all subsequent tables where they appear, quotes used as coding examples are shown verbatim, just as they appear in the transcripts, (e.g., “It’s a calendar they want want [*sic*] you to use”).

---

to pin down the timing of some other event of interest. Also as noted above, FRs were expected to support this process – the FR training program included training on when and how to encourage respondents to use the calendar aid in this manner. Table 5.a.2 (next page) reveals with quite distressing clarity just how infrequently FRs followed these training instructions. After the initial presentation of the calendar aid – or after the point in the interview when it *should* have been presented – the aid was only mentioned by either party in about ten percent of the interviews. Most of those cases were mentions only, with no suggestion by the FR that the respondent make note of some specific event or circumstance on the calendar. In fact, the most common “mention” of the calendar aid came in the form of a respondent’s request for clarification immediately after the FR’s introduction of it, which was typically followed by some form of expansion on the introduction by the FR. The coding process identified only five cases in which the FR encouraged the respondent to take the desired type of action.

Of course, respondents might have made entries on the calendar without it being evident in the recordings – that is, without any overt talk about it. Field test procedures discouraged FRs from collecting the calendars at the end of the interview (because possession of them would have raised data security issues), so there are no concrete data to examine. It seems unlikely, however, that this happened to any substantial extent. According to a post-interview debriefing question (administered by the FR to whoever was the last person in the household to provide information), about one-quarter of respondents said “yes” when asked (CAL\_2): “Did you write anything down on the calendar, to help keep track of events, or to remember when certain events

**Table 5.a.2: After Its Introduction, Was the Calendar Aid Ever Mentioned Again During the Interview?**

	<u>N</u>	<u>%</u>
1 – yes, with suggested action	5	4
2 – yes (mentioned only; no action)	10	7
<u>3 – no (includes omitted intro)</u>	<u>123</u>	<u>89</u>
TOTAL:	138	100%

Coding Examples:

- 1 – FR: [landmark events introduction]  
 R: I have, uh, I developed [type of cancer]  
 FR: Oooh!  
 R: And I was, um, that was – I found out at the end of [month1] I was, um, I believe my surgery was ... the beginning of [month2].  
 FR: Goodness, that’s, that’s definitely an event. ... So you can jot that down there, if you like, to help you remember things as we go along.
- 2 – FR: Well, ok ... you started – took a class last year. What time?  
 R: um [sigh] it must have been spring [pause] spring...  
 FR: ... So, spring, that would be – look at your little calendar, we got spring. Would you consider spring March, April, May?

occurred?” But the question was only asked of respondents to whom the FR reported having given a calendar, which was the case only a little over half of the time. When asked to rate on a five-point scale how helpful the calendar had been, where 1 was “very helpful” and 5 was “not helpful at all,” respondents were about four times more likely to pick a point on the “not helpful” side of the midpoint compared to the “helpful” side. Anecdotal reports from FRs also support the conclusion from the recordings that the calendar aid played no important role in the EHC interview. Such anecdotal evidence led to the decision to drop the calendar aid from SIPP-EHC interviewing procedures following the 2010 test.

5.b. Landmark events – introduction and probing

Borrowing from the psychology of autobiographical memory, EHC interviewing practices have generally included an opening question about “landmark events” – highly salient and memorable events in the respondent’s life that occurred during the survey’s reference period, and to which he or she can attach a date with great confidence. These events can be completely unrelated to the survey’s substantive goals, or they can be of direct interest; typically, there are no restrictions on content. All that matters is that the events (and especially their timing) be memorable to the respondent. The purpose of the landmark events is to provide memory anchors to assist the dating of other events and circumstances that *are* of substantive interest. Landmarks are a tool for interviewers; if a respondent cannot recall whether she started a job in March or April, for example, an interviewer might say, “Well, your brother’s thirtieth birthday party was in March,

you said. Did you have the job then, or did you start after that?” There is no need, of course, for cues of that sort to be limited to using landmark events – anything the respondent reports with confidence, including other events of substantive interest to the survey, can serve in that capacity. Because respondents produce landmark events from memory on their own, without the benefit of a targeted question, memory theory assumes they are recalled with great accuracy. Their main distinguishing feature in the EHC context, however, is simply that they provide the first opportunity to capture potentially useful memory cues, in case trouble should arise in the early substantive interview domains. And, practically speaking, their capture also increases the odds that useful cues will be available throughout the reference period of the interview.

The 2010 field test questionnaire included a scripted question to elicit landmark event reports<sup>12</sup>, and a set of circumstances, clearly defined in training, under which FRs were to probe for additional landmarks – or for any landmarks, in case none were forthcoming immediately. In brief, FRs’ training instructed them to try to capture two or three landmarks, including one from early in the previous calendar year (the survey’s reference period), but to do so “gently” and without spending substantial time in the effort. If the opening question and a mild probe or two failed to elicit more than one event, or failed to elicit any early-in-the-year events – or elicited none at all – the procedures called for the FR simply to drop the matter and move on to the substantive domains of the interview. This section of the report examines FRs’ performance in the 2010 field test with regard to introducing the landmark events “task” to respondents and probing for more (or any) landmarks according to the guidelines established in training.

Table 5.b.1 (next page) summarizes the results of the coding of the quality of FRs’ introduction of the landmark events reporting task. The coding scheme used here mirrors the one used in coding the presentation of the calendar aid. “Clear and coherent” (code 1) presentations were those which either followed the suggested scripted text, or touched on all of its key points (time period, memorability, sample contexts), using simple and direct language. Neither clear nor coherent introductions (3) were those I judged to have failed to deliver on those key points. I applied the middle, “ok/good” code (2) to introductions which fell short of the highest standard, but which still seemed to communicate some essential information.

For the most part, FRs carried out this part of the EHC interview quite well. As shown in the table, in approximately half (47%) of the recorded interviews the FRs presented the landmarks task in a manner judged to be clear and coherent. In most of these cases the FRs simply read the scripted introduction essentially verbatim (as is the case for the code 1 example included in the table) or followed it very closely. In another one-third (32%) of the cases I assigned the “ok/good” code, meaning that in about 80 percent of the recorded interviews FRs performed this task at the ok/good level or better.

---

<sup>12</sup> The scripted landmarks introduction, which followed immediately after a general description of the major topic areas to be covered in the survey, was as follows: “First, though, are there any events that occurred between January and December 2009, that stand out in your mind that you are able to date exactly or approximately? These might be family, financial, job, social or health related events. (Some examples include births or deaths, divorces or marriages, a vacation, health-related events such as an accident, major purchases that you have made, a job promotion or pay raise, a residence or job change.)”

---

**Table 5.b.1: Did the FR Present the Landmark Events Task in a Clear, Coherent Manner?**

	<u>N</u>	<u>%</u>
1 – yes, clear and coherent	65	47
2 – ok/good (not actively bad)	44	32
3 – no, not clear or coherent	17	12
7 – no, FR omitted landmark intro	11	8
(missing)	1	1
TOTAL:	138	100%

Coding Examples:

- 1 – FR: In this section of the interview we are interested in events that have occurred during the last calendar year, that is from January to December of last year. ... First, though, are there any events during the last calendar year, that is, from January to December of last year, 2009, that stand out in your mind that you're able to date exactly or approximately? These might be family, financial, job, social, or health-related events...
- 2 – FR: This section we interview with interests in events that occurred during the last calendar year, and that we said was from January to December 2009. I'm reading here. ... First, though, are there any events from the last calendar year, that is, from January to December of last year, that stands out in your mind that you are able to date exactly or approximately?
- 3 – FR: Were there any, like, big things that happened last year, any, like, milestone events? Like, you know, like, birth of a child?  
R: A what?  
FR: You know, like the birth of a child or a change of any circumstances.
- 

Table 5.b.2 (next page) summarizes respondents' initial responses following FRs' introduction of the landmark events reporting task. As shown in the first row, the most desired outcome – reporting a landmark event – occurred in about a third (31%) of the recorded interviews. In another one-quarter of the cases the respondent's first response was to indicate, in one way or another, that he or she had no landmark events to report – a perfectly legitimate response, but obviously not as desirable as reporting a landmark event. About one-third of the time respondents' first response was clearly not sufficient, either by indicating that they needed some additional clarification of the task set before them, or by responding in a variety of other ways that did not meet the objectives of the question. Finally, in about ten percent of the cases there either was no initial response (a result of the FR having failed to introduce the topic), or no *codeable* initial response.

Respondents' initial responses to the landmark events request proved to be significantly related, in predictable ways, to the quality of FRs' landmark introductions – see the results summarized in Table 5.b.3 (page 22). Clear and coherent (1) and ok/good (2) introductions of the landmark events task produced very similar distributions of initial responses (chi-square=2.5, 3 df, n.s.). Each of those distributions, however, differed significantly from the distribution elicited by neither clear nor coherent (3) introductions (1 vs. 3: chi-square=11.5, 3 df, p<.01; 2 vs. 3: chi-

---

**Table 5.b.2: What was the respondent’s initial response to the landmark events request?**

	<u>N</u>	<u>%</u>
1 – provided LM event	43	31
2 – “nothing happened”	34	25
3 – requested/signaled a need for clarification	22	16
4 – other response	24	17
7 – n/a, FR omitted LM intro	11	8
8 – missing, inaudible	3	2
<u>9 – missing, (LM interaction not recorded)</u>	<u>1</u>	<u>1</u>
TOTAL:	138	100%

Coding Examples:

1 – FR: [LM intro]

R: uh, my job, my [inaudible] my last job...

FR: Ok, so losing a job?

R: Yeah

2 – FR: [LM intro]

R: mmmm, [inaudible] hasn’t changed, life events hasn’t changed, school has not changed, and [inaudible] has not changed ... no, nothing has changed

3 – FR: So these might be like family, financial, job, social, um, birth, death, vacations, health-related, when did you start a job, when did you start college – anything that might, uh, stand out in your mind that you can pinpoint.

R: mmmm, when did I start college?

4 – FR: First, though, are there any events in the last calendar year, from January to December, that stand out in your mind that you’re able to date exactly or approximately?

R: Yes.

---

square=7.6, 3 df,  $p < .06$ ). Detailed comparisons indicate that ok/good or better introductions of the landmark task were significantly more likely to be associated with a desirable (or at least reasonable) response (code 1 or 2), and less likely to be associated with a problematic response (code 3 or 4), compared to introductions judged to be neither clear nor coherent. The differences were often of dramatic magnitude. For example, about one-third of the respondents who received an introduction that was at least ok/good provided a problematic initial response; among respondents for whom the FR’s introduction was of poor quality fully three-quarters produced a problematic initial response. Although these data are not sufficient for assessing a causal hypothesis, the results are certainly consistent with the notion that the quality of FRs’ performance matters, and that better performance elicits better outcomes from respondents.

As noted above, FR training stressed the goal of eliciting two or three landmark events from each respondent – FRs were to aim for that target, but not to press the issue. That is, if “gentle” probing failed to elicit the desired number of landmarks – or failed to elicit any at all – the FR

**Table 5.b.3: Respondents' Initial Response to the Landmark Events Question by the Quality of the FR's Introduction of the Task (Excluding Missing Data)**

	Quality of FR's Introduction of Landmark Events					
	1 – clear and coherent		2 – ok/good		3 – not clear or coherent	
	N	%	N	%	N	%
1 – provided LM event	28	44 <sup>a</sup>	13	30	2	13 <sup>a</sup>
2 – “nothing happened”	17	27	15	35 <sup>b</sup>	2	13 <sup>b</sup>
3 – req. clarification	8	13 <sup>c</sup>	8	19	6	38 <sup>c</sup>
4 – other response	11	17 <sup>d</sup>	7	16 <sup>c</sup>	6	38 <sup>d,e</sup>
TOTAL	64	100%	43	100%	16	100%
Codes 3 and 4 combined:	19	30% <sup>f</sup>	15	35% <sup>g</sup>	12	75% <sup>f,g</sup>

Statistical analysis summary (significant differences):

<sup>a</sup> t=2.36, 78 df, p<.05

<sup>c</sup> t=1.76, 57 df, p<.10

<sup>b</sup> t=1.70, 57 df, p<.10

<sup>f</sup> t=3.54, 78 df, p<.001

<sup>c</sup> t=2.41, 78 df, p<.05

<sup>g</sup> t=2.89, 57 df, p<.01

<sup>d</sup> t=1.79, 78 df, p<.10

was to drop landmarks and move on to the substantive EHC domains. But unless the respondent reported a number of landmarks in immediate response to the FR's presentation of the task, the FR was supposed to probe for more.

Table 5.b.4 (next page) summarizes FRs' probe-for-more-landmarks performance. In about 40 percent of the recorded interviews the FR probed for more (or any) landmarks in a manner that was in line with the goals of the interview and as the training program had intended – that is, they probed a little, but then quickly moved on to the main part of the interview. Adding in those whose probing performance I judged to be ok/good brings the positive performance figure to approximately 60 percent. Another small percentage of FRs also apparently followed what they had been instructed in training and correctly did *not* probe for more landmarks. In each of these cases, the respondent produced a good number of landmark events in response to the introductory question, obviating the need for any additional prompting. As the table clearly shows, actively poor performance on the part of the FR was a rare event – FRs' probes were either not clear/coherent (code 4) or pressed the matter excessively (code 5) in only about five percent of the interviews. (The latter error – excessive probing – was a particular concern of some members of the Research Group; the fact that that problem appeared in only one interview should allay those concerns.) Passively poor performance, however, is quite another story. In about one-quarter of the interviews, the FR failed to probe for more landmarks when the circumstances clearly called for such probing.

**Table 5.b.4: Did the FR Probe For More (or Any) Landmarks in a Clear, Coherent, and Appropriately Forceful Manner?**

	<u>N</u>	<u>%</u>
1 – yes, clear and coherent	56	41
2 – ok/good (not actively bad)	23	17
3 – correct no-probe (more probing unnecessary)	5	4
4 – no, not clear or coherent	5	4
5 – no, excessive/unnecessary probing	1	1
6 – no, failed to probe for more (in error)	36	26
7 – n/a, FR skipped landmarks	11	8
<u>missing (LM interaction not recorded)</u>	<u>1</u>	<u>1</u>
TOTAL:	138	100%

Coding Examples:

1 – [following the LM introduction, R has reported one major landmark event]

FR: Ok, anything else, umm, significant happen?

2 – [following the LM introduction, R has trouble coming up with any landmarks]

FR: Yes, yes, just anything that happened during that year. ... Nothing sticks out?

4 – [following the LM introduction, R has failed to produce any landmarks]

FR: Any? Ok, what happened on his work?

Field test procedures assumed that EHC-relevant events from early in the reference year would be most subject to recall difficulty. Thus, FR training stressed the importance of eliciting early-in-the-year landmarks from respondents, and of probing specifically for them, if necessary. Such probing was unnecessary if the respondent produced a sufficient number of landmarks with no (or minimal) probing of any kind, or if they produced an early-in-the-year landmark on their own, without any specific probing. Otherwise, however, the training instructed FRs to probe specifically for landmark events that occurred early in the reference year. As Table 5.b.5 (next page) shows, FRs mostly ignored or forgot about this facet of their EHC performance.

FRs probed for an early-in-the-year landmark event in only 8 of the 138 recorded interviews. The good news is that when they did so, they did so quite well – most of the probing was of high quality, all of the observed instances were judged to be at least ok/good, and there were no cases in which the probing for early-in-the-year landmarks was of actively poor quality (see codes 4 and 5). With regard to this particular type of probing, however, what FRs usually did was nothing. In about a quarter of the interviews this was the correct action – the respondent had already produced the desired total number of landmarks (I set three as the criterion), or had, without any specific prompting, reported at least one in the early part of the reference year (March or earlier). In 57 percent of the interviews, however, probing for an early landmark was called for but none was forthcoming from the FR. Restricting the focus to just those cases in which an early-in-the-year probe was clearly appropriate (i.e., excluding cases coded 3, 7, or

**Table 5.b.5: Did the FR Probe for Early-in-the-Year Landmarks in a Clear, Coherent, and Appropriately Forceful Manner?**

	<u>N</u>	<u>%</u>
1 – yes, clear and coherent	6	4
2 – ok/good (not actively bad)	2	1
3 – correct no-probe (early probing unnecessary)	39	28
4 – no, not clear or coherent	0	0
5 – no, excessive/unnecessary early probing	0	0
6 – no, failed to probe for early LMs (in error)	79	57
7 – n/a, FR skipped landmarks	11	8
<u>missing (LM interaction not recorded)</u>	<u>1</u>	<u>1</u>
TOTAL:	138	100%

Coding Examples:

1 – [the respondent has reported two landmark events, both in the fall]  
FR: Anything happen toward the first of the year last year?

2 – [the respondent has reported two landmark events in the same mid-year month]  
FR: Ok, umm, anybody [sic] earlier in the year or later?

missing) reveals in an even more negative picture. When the situation called for FRs to probe for an early-in-the-year landmark event, they failed to do so about 90 percent of the time ( $79/(6+2+79)=.91$ ). Further evidence suggests that the failure to probe was, in fact, a missed opportunity. Two of the eight instances in which FRs did probe the respondent for an early landmark were successful at eliciting at least one.

In addition to coding the recorded interviews for specific types of landmark event probing, I also coded them for a variety of miscellaneous inappropriate probing behaviors. As shown in Table 5.b.6 (next page), such behaviors were relatively rare – in about two-thirds of the interviews, FRs displayed no inappropriate probing behaviors of any kind. One result in Table 5.b.6 does deserve some special attention, however – the tendency for FRs to re-ask respondents for a new set of landmarks in proxy interview situations. Often these requests were somewhat bizarre – for example, the FR asked the proxy to report events that were highly memorable to the person being proxied for. The FR training package included seemingly clear instructions to FRs not to ask for new landmarks in proxy interview situations. And it explained why – namely that the only relevant memory material was in the head of the person answering the questions, not the person being answered about, and that had already been identified in the self-response interview of the person now serving as a proxy. Clearly, training was ineffective at communicating this concept. The recordings included 22 proxy interviews in which the landmark introduction and landmark probing had been part of the proxy respondent’s own self-response interview; in 17 of those cases (77%) the FR improperly probed for new landmarks at the outset of the proxy interview. (Table 5.b.6 shows only 14; three others are include in the “other/misc.” category.)

---

**Table 5.b.6: Did the FR Probe Inappropriately for Landmarks in any (Other) Manner?**

	<u>N</u>	<u>%</u>
1 – no, no inappropriate probing	94	68
2 – yes, excessive probing	6	4
3 – yes, probed a proxy for new landmarks	14	10
4 – yes, leading probes for unnecessary landmarks	6	4
5 – yes, other/misc. (includes multiple codes)	7	5
<u>7 – n/a, FR skipped landmarks</u>	<u>11</u>	<u>8</u>
TOTAL:	138	100%

Coding Examples:

2 – [after fairly extensive probing by the FR, and with some difficulty, the respondent has reported one landmark event]

FR: Ok, now, anything else that you can think of? Did you get a new car?

R: Nope.

FR: Uhh, did anybody graduate? Move – well, not move out, did anybody graduate?

R: Did she graduate last year? Nah, I think [inaudible]

FR: Did anybody – did, well, [inaudible] not nothing like that, but did anybody die? Did you have to go in the hospital for anything? Anything health-related? Anything special at work, like a promotion? Anything like that?

R: Nothing.

4 – [the respondent has reported one landmark event]

FR: Ok, anything else that stands out in your mind that you may be able to date exactly, maybe later on in the year? The birthday of [the respondent’s very young twins]?

R: Yeah, [month] [day]

FR: Ok, that was a good landmark

5 – FR: Ok, now I am going to ask you for some landmark events such as, when did you move into this, ah, apartment?

---

### 5.c. Respondents’ landmark event reports

This section examines the outcome of all of the FRs’ introductions of, and probings for, landmark events, starting with Table 5.c.1 (next page), which shows the number of landmark events reported in the 138 recorded interviews. Respondents produced at least one landmark event report in about two-thirds of the recorded interviews; the maximum number observed was six. Across all interviews, respondents reported an average of 1.3 landmarks; respondents who reported at least one landmark reported about two, on average.

The number of landmark event reports affords the first opportunity to compare the measurement of a phenomenon in the recorded interviews with the same measurement from non-recorded interviews. The comparison reveals striking and provocative differences. The full data file for the 2010 field test contains 3,871 self-response, adult (age 15+), complete or “sufficient partial,”

---

**Table 5.c.1: How many landmark events did respondents report?**

	<u>N</u>	<u>%</u>
0 LMs	49	36
1	42	30
2	25	18
3	10	7
4	6	4
5	5	4
6	1	1
<u>TOTAL:</u>	138	100%

Total # of landmark events reported:	177
Mean # of landmarks, all interviews:	1.3
Mean # of landmarks, all interviews with 1+ LM:	2.0

---

English-language interviews<sup>13</sup> with a first-listed household member, a set of characteristics that is comparable to the circumstances of recorded interviews. In stark contrast to the recorded interview outcomes, a substantial majority – 58% – of the 3,733 otherwise comparable but non-recorded interviews in the 2010 field test yielded no landmark event reports. The difference between this figure and the proportion of recorded interviews which produced no landmark event reports (36%) is highly significant ( $t=5.23$ ,  $p<.0001$ ). This largely explains a dramatically lower average number of landmark reports per respondent in the larger sample – 0.7, versus 1.3 in recorded interviews ( $t=5.29$ , 141.76 df (adjusted for unequal variances),  $p<.0001$ ). But it doesn't explain all of the difference, because even those who reported one or more landmarks tended to report fewer in the larger, unrecorded sample (1.6) than in the recorded cases (2.0;  $t=3.24$ , 92.63 df (adjusted for unequal variances),  $p<.005$ ).

The differences in landmark event production between recorded and unrecorded interviews are not due to some form of bias in the selection of FRs for recording duty. This is apparent in Table 5.c.2 (next page), which starts with the subset of comparable interviews from the full data file (as described above), but adds the further restriction of only looking at the unrecorded interviews conducted by FRs who recorded. Even within a constant set of FRs, huge differences remain in the “yield” of landmark event reports from recorded and unrecorded interviews.

Speculation about the cause of these large differences is easy; concrete proof is another matter. One possibility that immediately springs to mind, of course, is the notion that when they were conducting recorded interviews, FRs took greater pains to conduct them “by the book” than they did when conducting unrecorded interviews. Another derives from the fact that the 2010

---

<sup>13</sup> Early in the 2010 interview, the variable HHSPEAK1 asked FRs to report whether any part of the interview up to that point had been conducted in a language other than English. For this analysis, I excluded all otherwise-qualified interviews in which FRs entered a “yes” response to HHSPEAK1.

**Table 5.c.2: Landmark Event Reporting in Recorded Interviews and in All Other Comparable Interviews [see text] Conducted by FRs Who Recorded**

	<u>Recorded Interviews</u>	<u>All Other Comparable (non-recorded) Interviews by FRs Who Recorded</u>
N (%) who reported 0 LMs	49 (36%) <sup>a</sup>	520 (58%) <sup>a</sup>
N (%) who reported 1+ LM	89 (64%)	381 (42%)
<b>TOTAL</b>	<b>138 (100%)</b>	<b>901 (100%)</b>
Mean # of LMs, all interviews:	1.3 <sup>b</sup>	0.6 <sup>b</sup>
Mean # of LMs, interviews with 1+ LM:	2.0 <sup>c</sup>	1.5 <sup>c</sup>

Statistical analysis summary (significant differences):

<sup>a</sup> t=4.93, 1037 df, p<.0001

<sup>b</sup> t=5.40, 154.53 df (adjusted for unequal variances), p<.0001

<sup>c</sup> t=3.67, 103.56 df (adjusted for unequal variances), p<.0005

procedures did not exert tight control over the selection of cases for recording. Perhaps FRs' autonomy led them, consciously or not, to skew those selection processes toward respondents of a certain type – more amenable, more voluble, etc. The data are insufficient to shed much light on how – or even whether – these or other such biases might affect the results of this investigation. One possible impact might be to paint an unrealistically positive picture of FRs' administration of the 2010 EHC interview. As will be clear later (see section 7.b), the recordings suggest that when FRs had the opportunity, they did not carry out key EHC interviewing procedures particularly well. So, even if the findings are biased in an overly positive direction, the bottom-line conclusion – that better FR training is needed – remains unchanged.

Again – one can speculate, but identifying the causes of a higher yield of landmark events from recorded interviews is beyond the scope of the present research, and the impacts of that phenomenon on its results and conclusions remain unclear. In the analyses to follow I simply work with the results as they emerged from the recordings, while trying to remain alert to possible influences of “performance bias” on the conclusions from those results.

Table 5.c.3 (next page) addresses the relationship between the overall quality of FRs' presentation of the landmark task and respondents' production of landmark events. As an operational definition of the “overall quality” of FRs' performance I added together two measures: initial introduction of landmarks and subsequent probing for more, assigning each measure a score of 2 for “clear and coherent” performance, 1 for “ok/good” performance, and 0 for performance judged to be neither clear nor coherent. The resulting overall quality scale thus ranges from 0 (poor performance on both components) to 4 (“clear and coherent” performance on both components); the analysis excludes cases in which either component was missing.

**Table 5.c.3: Number of Landmarks Reported and the Overall Quality of FRs' Presentation of the Landmark Events Task**

# of LMs Reported	Overall Quality Score [see text] – FRs' Landmark Events Presentation									
	0		1		2		3		4	
	N	%	N	%	N	%	N	%	N	%
0	5	42	11	61	12	44	6	21	4	10
1 or more	7	58	7	39	15	56	23	79	36	90
TOTAL:	12	100%	18	100%	27	100%	29	100%	40	100%
<u>Mean # of LMs</u>										
all Rs:	0.8		0.6		1.0		1.6		2.1	
Rs w/ 1+ LM:	1.4		1.6		1.7		2.0		2.3	

Statistical analysis summary (significant differences):  
[see text]

Statistical tests confirm, in various ways, a clear, positive relationship between the overall quality of the FR's presentation of the landmark task (as defined above) and the number of landmarks produced by the respondent. A simple test of association between those two variables yields a positive and highly significant Pearson correlation coefficient ( $r=+.37$ ,  $p<.0001$ ). Additional tests comparing the extreme ends of the scale – collapsing the 0 and 1 overall quality scores into a single “low quality” category, and, similarly, the 3 and 4 scores into a “high quality” category – present a consistent story. The mean number of landmarks produced following high quality presentations (1.9) is significantly greater than the corresponding low quality mean (0.7,  $t=4.11$ , 97 df,  $p<.0001$ ); as is the proportion of respondents who reported at least one landmark (86% vs. 47%,  $t=4.37$ , 97 df,  $p<.0001$ ); as is the mean number of landmarks among respondents who produced at least one (2.2 vs. 1.5,  $t=1.89$ , 71 df,  $p<.10$ ).

What kinds of events did respondents report as landmarks? Table 5.c.4 (next page) summarizes the subject matter of respondents' landmark events, broadly categorized according to whether or not the landmark was a domain event. As noted earlier, the role of the landmarks – as memory anchors to assist recall, later on in the interview, of the timing of events of substantive interest to the survey – is not unique to them. Any event or circumstance recalled confidently can serve as a memory anchor for later material. Thus, to the extent that respondents' landmark events deal with subject matter soon to be covered in the main body of the interview – job starts and stops, for example, or changes in residence – it is arguably less necessary to devote time at the beginning of the interview trying to elicit them as a special category of events. One should, then, just plunge into the interview and use the events uncovered in its earlier portions to assist recall of events dealt with later in the interview, as necessary. As shown in the table, about a third of respondents' landmark event reports involved domain matters that they probably would have reported in the normal course of the SIPP-EHC interview without any special probing; the most

**Table 5.c.4: Subject Matter of Reported Landmark Events**

		<u>N</u>	<u>%</u>
EHC domain LM events involving household members:	Residence	14	8
	Education	13	7
	Employment	25	14
	Programs; health insurance	7	4
	<u>Marital status</u>	<u>2</u>	<u>1</u>
	Domain Sub-Total:	61	34
LM events outside the EHC domains (or involving non-hh members) [*]	Life course event	21	12
	Health/trauma/death	32	18
	Occasions	43	24
	Major financial event	3	2
	Calendar event	0	-
	<u>Other, misc.</u>	<u>16</u>	<u>9</u>
		Non-Domain Sub-Total:	115
	<i>(missing/uncodeable</i>	<i>1</i>	<i>1)</i>

\* Life course event (engagement, pregnancy, citizenship, baptism, bar/bat mitzvah, etc.); Health/trauma/death (illness, medical procedure, accident, injury, crime, etc.); Occasions (travel, vacation, anniversary, visit, etc.); Major financial event (income change, major purchase, foreclosure, “money trouble,” etc.); Calendar event (Memorial Day, July 4<sup>th</sup>, Thanksgiving, etc.); Other (domain category events involving non-members of the household, miscellaneous)

frequently cited domain involved employment-related events. Conversely, the great majority of the landmark events – about two-thirds of them – were unique to that interview topic, and would not have been available later for use as memory anchors had there been no special probing for them at the beginning of the interview. Hillmer, Monte, and Fields (2011) report similar findings in an analysis of all 2010 field test interviews.

Another interesting facet of respondents’ landmark reports is their positive or negative character. During the development of the 2010 field test some members of the Research Group (and others) expressed concern that respondents might perceive the request for landmarks as pressure to dredge up unpleasant memories. For each reported landmark event I coded whether it was clearly positive, clearly negative, somewhere in between (either neutral or a mixture of positive and negative), or uncertain; the results are summarized in Table 5.c.5 (next page).

As is often the case in studies of human behavior (see, e.g., Matlin and Stang, 1978; Armor and Taylor, 2002), there is a clear tendency toward positivity in respondents’ landmark event reports, with about half of the reported events describing clearly positive incidents. On the other hand, there is little evidence of any substantial reluctance to mention negative events – deaths,

---

**Table 5.c.5: Positive/Negative Character of Reported Landmark Events**

	<u>N</u>	<u>%</u>
1 – Positive	91	51
2 – Neutral; both + and -	5	3
3 – Negative	51	29
4 – Uncertain; can't tell	30	17
<b>TOTAL:</b>	<b>177</b>	<b>100%</b>

Coding Examples:

1 – R: I had a granddaughter

2 – R: My daughter had her third little boy [born with Down Syndrome]

3 – R: My mother expired in [month]

4 – R: I went to visit my mom and stuff

---

illnesses, job losses, break-ups, thefts, accidents, and the like featured prominently in respondents' reports of important events in their lives in 2009, comprising about one-third of all landmark reports. Although it might not have been revealed in respondents' overt speech, in no transcript did I detect any concern that they felt pressured to focus on painful or unpleasant memories – or, for that matter, pressured to avoid them. Rather, the sense of the interaction was much more mundane: bad things happen to everyone at some time – that's life. It is perfectly natural that, if a survey asks respondents to report important, highly memorable events in their lives, quite negative stories will occasionally emerge<sup>14</sup>.

The final characteristic of respondents' landmark event reports that I designed my coding scheme to capture was their timing – specifically, the month in which each event occurred. Recall that the training instructed FRs to focus some energy on eliciting landmarks from early in the reference year, if possible. The motivation for this was the assumption that respondents' recall difficulties might congregate there, at the furthest distance in time from the date of the interview, and thus that memory anchors for that part of the year might be of most use. We have already seen, in section 5.b (see Table 5.b.5), that FRs did not perform particularly well at this aspect of their landmark events duties – for the most part they simply did not probe at all for early-in-the-year landmarks. The data certainly bear that out in respondents' behavior as well, offering no evidence that landmark event reports skewed toward the early months of the

---

<sup>14</sup> The 2010 training package took pains to alert FRs to the likelihood that some landmarks would be so negative as to require them to step out of the interviewer role briefly to express basic human sympathy. The transcripts reveal that some FRs handled this quite well (“Your mom died in [month]? Sorry to hear that.”) Too often, unfortunately, FRs appeared to be so caught up in the immediate task that they failed to do the right, one-human-being-to-another thing. For example: in response to a respondent's report that in the same month one of her best friends died and her son was shot and killed, the FR said, “OK; anything else?” Moore (2011b) cites several other examples.

reference year – see Table 5.c.6. A chi-square test on these results (chi-square=6.28, 11 df, n.s.) fails to reject the null hypothesis of an even distribution of landmark events across all months of the year.

---

**Table 5.c.6: Timing (Month) of Reported Landmark Events**

	<u>N</u>	<u>%</u>
Jan	14	8
Feb	15	8
Mar	8	5
Apr	16	9
May	14	8
Jun	15	8
Jul	12	7
Aug	18	10
Sep	18	10
Oct	14	8
Nov	14	8
Dec	11	6
<u>(missing)</u>	<u>8</u>	<u>5</u>
TOTAL:	177	100%

Statistical analysis summary (significant differences):  
[none]

---

#### 5.d The concordance of respondents’ landmark event reports and FRs’ data entries

The recorded interviews allow an evaluation of the quality of FRs’ performance with regard to whether they captured respondents’ landmark event reports accurately. The general question here is: Is what we see in the data an accurate representation of what respondents said in the interview? Not that there has been concern on this matter – rather, the recordings simply present an unusual opportunity to examine this link of the deductive chain. Survey analysts generally assume that survey data offer direct access to respondents’ interview reports, when in fact, in interviewer-administered surveys, that access is indirect, with the FR serving as a mediator. Examining the underlying assumption may be particularly important when the main point of the analysis is to address methodological issues.

This section examines two very general indicators of the concordance between the words that respondents actually spoke in response to the landmark events reporting task and the data that FRs entered into the questionnaire. The first of those indicators is simply the number of landmark events. On this dimension FRs performed very well; discrepancies between the number of reported landmarks and the number captured in the data file are rare. In 83 percent of

the interviews the number of landmarks reported by the respondent matched exactly the number recorded by the FR; 99 percent of the time the discrepancy was no greater than one – see Table 5.d.1. The data suggest a tendency for FRs to “miss” a reported landmark ((16+2)/138=13%) more often than they recorded one in error (5/138=4%;  $t=2.83$ , 136 df,  $p<.01$ ), but the “big picture” result is strong support for the conclusion that FRs’ data file entries represent quite accurately the number of landmarks reported by respondents.

**Table 5.d.1: Did FRs Record the Same Number of Landmarks as Respondents Reported?**

	<u>N</u>	<u>%</u>
D = 0 (no discrepancy; data file total = transcript total)	115	83
D = -1 (transcript total < data file total)	5	4
D = +1 (transcript total > data file total)	16	12
<u>D = +2 (transcript total &gt; data file total)</u>	<u>2</u>	<u>1</u>
TOTAL:	138	100%

The same picture emerges from an examination of the content of FRs’ landmark event entries. As shown in Table 5.d.2, in almost all cases the questionnaire entry closely matched the respondent’s report. In most cases, in fact, the questionnaire entry was almost a word-for-word match with what respondents said, as in the code 1 example shown in the table. Again, these

**Table 5.d.2: Did FRs’ Landmark Event Description Entries Match Respondents’ Reports?**

	<u>N</u>	<u>%</u>
1 – yes	149	95
2 – uncertain	4	3
<u>3 – no</u>	<u>4</u>	<u>3</u>
TOTAL:	157[*]	100%

\* This table excludes 20 landmarks which were reported by respondents but which did not appear in the data file (see the “D = +1” and “D = +2” rows of Table 5.d.1).

Coding Examples:

- 1 – R: Well, I was dismissed from graduate school  
[FR’s questionnaire entry: dismissed from grad school]
  
- 2 – R: My son got deported  
[FR’s questionnaire entry: son depoyed *[sic]*]
  
- 3 – R: In *[month]* he got laid off  
[FR’s questionnaire entry: laid off of work *[recorded in a different month]*]

data offer strong support for confidence in FRs' performance with regard to capturing respondents' landmark event reports accurately.

## 6. Results III: FR Characteristics and EHC Performance

### 6.a. Relationship between certification test score and EHC performance

I appended FRs' certification test results for the SIPP sub-scale, the EHC sub-scale, and the total test score to the recorded interview data file. To simplify an analysis of the relationship between test scores and interview performance, I recoded each of the individual EHC performance variables – calendar aid introduction, landmark introduction, landmark probing, etc. – into a simple, dichotomous, good performance/bad performance score<sup>15</sup>. Table 6.a.1 (next page) presents and compares the test score means for “good” performing and “bad” performing FRs on each of five individual performance measures. I also carried out a test of the association between the test scores and the overall quality of FRs' landmark events presentations graded on a 0-4 scale (see section 5.c); the table also shows the relevant Pearson correlation coefficients. (The numbers of cases displayed in each set of rows refers to the number of interviews (out of 138) in which the FR's performance qualified as “good” or “bad,” and in which his/her certification test score was non-missing. Because of this arrangement, FRs who conducted more than one recorded interview contribute their test data multiple times to the results for each performance measure – quite possibly contributing to the mean test score estimate in both the “good” and “bad” category for the same performance measure.)

Obviously, all of the performance measures summarized in the table focus on EHC-related interviewing procedures; thus, if there were any relationship between test score and EHC performance we should expect to see it more strongly for the EHC sub-scale than for the SIPP sub-scale. And, indeed, this is what the results strongly suggest. None of the performance measures shows a significant relationship with scores on the SIPP sub-scale of the certification test. This is in marked contrast to the story for the EHC sub-scale, where the relationship of test score and EHC interviewing performance is statistically significant in the expected direction for four of the six performance measures. In two of those cases the total score on the certification test is also significantly related to performance, but it is quite clear that the EHC sub-scale is the main “driver” of the test-performance relationship.

### 6.b. Relationship between FR experience/tenure and EHC performance

Table 6.b.1 (page 35) summarizes, for each of the five individual performance measures, the percentage of FRs in each experience/tenure group whose performance (recoded dichotomously as described above) was judged to be “good;” for the overall landmark presentation quality scale the table shows mean performance scores (on the 0-4 scale) across the three experience/tenure groups. (The (n) in each table cell indicates the number of recorded interviews (out of 138) conducted by an FR with that experience/tenure level and with a non-missing code for the

---

<sup>15</sup> See sections 5.a and 5.b for details regarding the individual performance variables. In general, the “good” performance category combines the “clear and coherent” and “ok/good” codes; “bad” performance combines the “not clear or coherent” code with those who omitted the task entirely.

**Table 6.a.1: Certification Test Score Means and EHC Performance**

<u>Mean Certification Test Scores and ...</u>		<u>SIPP</u> <u>Sub-scale</u>	<u>EHC</u> <u>Sub-scale</u>	<u>Cert Test</u> <u>TOTAL</u>
a) ... the Quality of FRs' Introduction of the Calendar Aid	“good” (n=91) “bad” (n=19)	19.9 19.7	13.5 <sup>a</sup> 11.9 <sup>a</sup>	33.5 31.6
b) ... the Quality of FRs' Introduction of Landmarks	“good” (n=90) “bad” (n=23)	20.0 19.5	13.6 <sup>b</sup> 11.7 <sup>b</sup>	33.6 <sup>c</sup> 31.3 <sup>c</sup>
c) ... the Quality of FRs' Probing for More Landmarks	“good” (n=69) “bad” (n=36)	20.0 19.7	13.7 <sup>d</sup> 12.6 <sup>d</sup>	33.8 32.3
d) ... the Quality of FRs' Probing for Early Landmarks	“good” (n=67) “bad” (n=38)	20.2 19.8	13.8 13.1	33.9 32.9
e) ... Bad probing (in general) for Landmarks	“good” (n=79) “bad” (n=27)	19.8 20.3	13.3 13.3	33.1 33.6

Statistical analysis summary (significant differences):

<sup>a</sup> t=2.66, 108 df, p<.01      <sup>c</sup> t=2.27, 111 df, p<.05  
<sup>b</sup> t=3.52, 111 df, p<.001      <sup>d</sup> t=2.30, 103 df, p<.05

f) Correlation (r) Between the Overall Quality of FRs' Presentation of the Landmark Events Task (see section 5.c) and Certification Test Sub-scale and Total Scores

<u>SIPP</u> <u>Sub-scale</u>	<u>EHC</u> <u>Sub-scale</u>	<u>Cert Test</u> <u>TOTAL</u>
+ .11	+ .28 <sup>e</sup>	+ .22 <sup>f</sup>

Statistical analysis summary (significant correlations):

<sup>e</sup> 105 df, p<.005      <sup>f</sup> 105 df, p<.05

performance measure of interest. As in the previous table, FRs who recorded multiple interviews are represented in the results multiple times for each performance measure.)

Looked at individually, the results in Table 6.b.1 appear at first to be unambiguous. None of the EHC performance variables shows any evidence of a significant relationship with experience/tenure; none of the 18 paired comparisons is statistically significant. Looked at more globally, however, interesting patterns emerge. Perhaps not surprisingly, the “new hire” group was in general the lowest-performing group; the rank ordering of each of their observed estimates is the lowest of the three FR experience/tenure groups on four of the six performance measures, and never ranks first. The experienced (non-SIPP) group presents the opposite pattern; their estimates rank first on four of the six performance measures, and never rank last. A simple sign

---

**Table 6.b.1: FR Experience/Tenure and EHC Performance**

<u>FR Experience/Tenure and ...</u>	<u>SIPP- Experienced</u>	<u>Other Experienced (non-SIPP)</u>	<u>New Hire</u>
a) ... the Quality of FRs' Introduction of the Calendar Aid – % “good” (n)	81.6 (49)	81.0 (42)	74.4 (43)
b) ... the Quality of FRs' introduction of Landmark Events – % “good” (n)	76.0 (50)	84.1 (44)	79.1 (43)
c) ... the Quality of FRs' probing for More Landmark Events – % “good” (n)	64.4 (45)	71.4 (42)	64.1 (43)
d) ... the Quality of FRs' Probing for Early-in-the-Year Landmark Events – % “good” (n)	37.8 (45)	42.9 (42)	30.8 (39)
e) ... Bad Probing (in general) for Landmark Events – % “no bad probing” (n)	67.4 (46)	78.6 (42)	76.9 (39)
f) ... the Overall Quality (0-4 scale) of FRs' Landmark Events Presentation – mean (n)	2.7 (45)	2.5 (42)	2.3 (43)

Statistical analysis summary (significant differences):  
[none, but see text]

---

test for those outcomes – no first place rankings or no last place rankings in six “trials” – finds that they are sufficiently unlikely ( $p < .10$ ) as to provide some support for rejecting the null hypothesis.

The notion that there might be performance differences according to FR experience/tenure makes some intuitive sense, particularly for the new hires. Developing interviewing proficiency takes time and practice, and the new hires had, presumably, very little interviewing experience to draw on; this was almost certainly the case relative to the other groups. Additionally, unlike the two experienced groups, new hires had had little chance to have their ranks “thinned” of very poor performers. That the experienced (non-SIPP) group was, overall, the best performing group also makes some sense, especially considering that some EHC interviewing practices differ in major ways from standard SIPP interviewing practices. If those differences were substantial enough, then SIPP-experienced FRs, relative to FRs with other types of experience, might start out at a slight disadvantage by having to spend time and effort “unlearning” old habits. Obviously, this is largely conjecture – and conjecture about marginal differences, at that.

## 7. Results IV: Using the Recordings to Identify Respondent Recall Difficulties and to Examine FRs' Responses to Them

Despite the great attention paid to them thus far, it is important to remember that landmarks are not actually of substantive interest. They are only a tool of the EHC trade, for use, as needed, in helping respondents recall the timing of events that *are* of substantive interest. Not only are landmark events “just” of instrumental, rather than substantive, value – they are not even of unique instrumental value, since any event recalled with confidence, including events from the survey’s substantive domains, can serve the same function. Landmarks may be particularly salient in memory, and they may have some special value because they can be deployed very early in the interview, before many domain events have been reported, but for the most part they simply add to the arsenal of potential memory anchors for FRs to call on as needed.

Regardless of their source, memory anchors are, at least in theory, vital to the quality of survey data for autobiographical events, and an essential component of the procedures that distinguish EHC-style interviewing from a conventional interviewer-administered interview. The strength of the EHC method rests on the notion that memory anchors, used appropriately, will assist respondent recall without biasing it. In a well-designed EHC questionnaire, the full array of possible memory anchors will be easily accessible to the interviewer, should the need for them arise. And an effective EHC training program will produce interviewers who are alert to signs of trouble, and whose response repertoires are fine-tuned to yield quick and effective action.

This final analysis section focuses on just those sorts of instrumental issues: How much demand was there for the “tricks-of-the-trade” of EHC interviewing? That is, was there evidence that respondents experienced any difficulty recalling the timing of EHC-related events? If so, did FRs notice? Did FRs take any action to help solve those problems? What sorts of action did they take, and what was the quality of those actions? Did it appear that the recall problems were resolved – either because of, or in spite of, or independent of the FRs’ actions?

### 7.a. Evidence of respondent recall problems

My review of the interview transcripts examined respondents’ verbal behaviors for evidence of difficulty recalling when some event happened – typically, the start or stop of a spell of some kind. These same respondent behaviors were supposed to cue FRs to take corrective, on-the-spot, EHC-style action to assist recall – the most favored form of which would be to use some salient event in the respondent’s life as a memory anchor (e.g.,: “You moved to this apartment in April, you said. Did your Food Stamps end before you moved here, or was it after?”). Those potential memory anchors primarily reside in the main, calendar portion of an EHC interview. The design of the instrument used in the 2010 field test made it difficult for the FR to access the calendar portion of the interview once it was completed. Because of this factor, I limited my coding of the presence of recall difficulties to the EHC calendar portion of the interview, where the tools for the FR to take corrective action were most immediately accessible. (This supposed coding limitation had minimal practical implications, since the vast majority of recall problem “signals” occurred during the administration of the EHC.) Table 7.a.1 (next page) summarizes the coding results.

**Table 7.a.1: How Often Did Respondents Show Evidence of a Problem Recalling the Timing of an EHC Event or Circumstance?**

	<u>N</u>	<u>%</u>
0 – no recall problems	117	85
1 – 1 recall problem	19	14
2 – 2 recall problems	1	1
3 – 3 recall problems	1	1
	138	100%

Total number of recall problem “signals”: 24

Coding Examples (respondent verbalizations coded as evidence of a recall problem):

FR: In which month in 2009 did you receive any special payments such as bonuses, awards, or large commissions or receive no special payments?

R: I was thinking there might have been, ah, award - cash award [inaudible]. I couldn't tell you [the month]

FR: Ok and in what month did she work for pay?

R: She worked ... the end of [month1], she started the end of [month1] ... in this job until, ah, I don't know, I'll say [month2]

FR: OK.

R: I guess.

Overt signs from the respondent that he or she was having trouble recalling the timing of an EHC-relevant event were rare. In fact, the vast majority of interviews – about 85 percent – proceeded without any such signals from the respondent. This is not to say that respondents in these interviews had no recall problems, or that they reported everything accurately, but simply that their overt behavior provided no reason for an FR to take any corrective action. Among the approximately 15 percent of interviews in which there was evidence of a recall problem, the vast majority had only one such incident. In total, across all 138 recorded and transcribed interviews, I counted 24 instances in which a respondent's verbal behavior gave evidence of an underlying problem in recalling the timing of some EHC-related event.

7.b. FRs' responses to evidence of respondent recall problems

Of course, a problem must first “register” with an FR in order for him or her to respond to it. As shown in Table 7.b.1 (next page), when respondents displayed signs of a recall problem FRs picked up on them only imperfectly – between one-quarter and one-third of the time there was no evidence in the FR's behavior that he or she noticed that there might be a problem.

FRs' own reports about respondents' recall problems offer an interesting contrast to the results of the coding of the transcripts. A debriefing question at the end of each completed interview asked the following question:

---

**Table 7.b.1: When the Respondent Signaled a Recall Problem, Did The FR Show Any Evidence of Noticing It?**

	<u>N</u>	<u>%</u>
1 – yes	17	71
2 – no	7	29
TOTAL:	24	100%

Coding Examples:

1 – FR: In which months of 2009 did you receive any special payments such as bonuses, awards or large commissions?

R: Umm, oh, couldn't really tell you on that one...

FR: You did receive... Do you remember what month at all that you received the bonus?

2 – FR: Ok, do you – now they want to know if you had any Food Stamps issued to you

R: Uh, the beginning of the year... yes, the beginning of the year, from [month1] to, like, [month2] or [month3].

FR: Ok.

---

“During the calendar portion(s) of this interview, did a respondent have trouble remembering when some spell started or stopped, or the date of any other event?”

FR debriefing data are missing for 5 of the 138 recorded interviews. In the 133 interviews with non-missing data FRs responded “yes” to this question 24 times, which, in both absolute numbers and percentage terms, compares favorably to the frequency of problems observed in the recorded interview transcripts. A direct, case-by-case comparison reveals, however, that this overall comparability is largely illusory, resulting from a substantial number of largely compensating discrepancies – see Table 7.b.2.

---

**Table 7.b.2: Were There Any Respondent Recall Problems? Matched Transcript Coding and FR Debriefing Results**

Did the Interview Include a Recall Problem According to...		... FR Debriefing?		
		<u>no</u>	<u>yes</u>	
<u>... Transcript Coding?</u>	<u>no</u>	96	16	
	<u>yes</u>	13	8	
		109	24	(TOTAL: 133)

---

The 16 discrepancies in the upper-right cell of Table 7.b.2 may have a benign explanation in the restricted focus of the transcript analysis. The debriefing questions asked each FR to reflect back on the entire interview for the entire household, whereas the transcript analysis focused on the EHC portion of the interview, and the behavior of a single respondent. It is entirely possible that no recall problems surfaced in the recorded interview but that they did appear in later interviews with other respondents in the household. The other type of discrepancy – the 13 cases (out of 21) in which the review of the transcript found evidence of a recall problem that the FR failed to report in the debriefing questions – is somewhat less conducive to easy explanation. The debriefing questions were well after-the-fact; perhaps by the time they appeared there was a tendency for the FR to forget a brief incident or two that occurred early on in a long interview. This could explain why the FRs’ real-time behaviors suggest that they recognized respondent recall problem incidents at a much higher rate (17/24, 71% – see Table 7.b.1) than is suggested by their after-the-fact memories for those events (8/21, 38%;  $t=2.20$ , 43 df,  $p<.05$ ).

When they noticed a recall problem, FRs almost always took some form of action to try to assist the respondent<sup>16</sup>. Of the 17 “noticed” recall problem incidents, FRs tried to assist in all but one instance, which proceeded as follows:

FR: In what months did you first start to look for work in 2009?

R: I don’t remember

FR: You don’t remember, I can put don’t know, that’s ok. All right, so you don’t remember like when you stopped looking for month, the first time, what month you stopped looking for work?

R: No...

FR: Ok. ... so you don’t remember like what the other time, what months?

R: No

FR: Ok, I’ll just put that in. I’m just answering don’t know...

This FR’s repeated questioning only sought confirmation of the respondent’s inability to recall when he/she looked for work. While this behavior is clear evidence that the FR noticed the problem, there was no assistance offered. No *positive* assistance, to be precise, directed toward advancing the goals of the survey – just an easy escape route.

When FRs took action to assist, what was the nature of that action? Table 7.b.3 (next page) addresses this question. In only one instance did an FR, faced with a respondent recall problem, call on an already-reported event (a landmark, in this case) to help the respondent pin down the timing of the event in question – the technique which constitutes the unique strength of EHC interviewing methods, and which was the focus of considerable attention in training. The interaction details for the lone instance of EHC-style probing show both the relative simplicity of the technique and its apparent effectiveness.

---

<sup>16</sup> “Noticing” a respondent’s recall problem and taking action to address that problem are conceptually distinct, but in practice – in a research project whose data are limited to observations of overt behavior – the two concepts proved very difficult to disentangle. In general, the FR’s action constituted the primary evidence that he/she had noticed the problem; conversely, the absence of any action was generally the primary evidence that he/she had failed to notice.

**Table 7.b.3: If the FR Offered Assistance to Address a Recall Problem, What Was the Nature of That Assistance?**

	<u>N</u>	<u>%</u>
1 – LM event (non-domain)	1	6
2 – domain event	0	–
3 – LM event (domain)	0	–
4 – general calendar anchor (holiday, season, etc.)	4	25
<u>5 – other assistance</u>	<u>11</u>	<u>69</u>
TOTAL:	16	100%

Coding Examples:

- 1 – R: I didn't work at all last year... Oh yes, I did, at the beginning of the year I did  
FR: You did? Okay... Let me see then that was-  
R: I believe it was ummm -  
FR: What month?  
R: I'm tryin' to think of - but yeah it was, [month1]... I know it was during [month1]  
FR: And for about how - until what month?  
R: Ummm, only worked for about mon- two months  
FR: So that was before – you, you ended your job before your son went to - got married?  
R: yeah  
FR: So, so maybe just the two months, [month1] and [month2]...  
R: ... yeah [month1] and [month2]
- 4 – FR: In which month in 2009 did you receive any special payments such as bonuses, awards, or large commissions or receive no special payments?  
R: I was thinking there might have been, ah, award - cash award [inaudible]. I couldn't tell you [the month]  
FR: Mmm, let's see - can you think of the weather at the time? Maybe that would help you to recall, 'cause I need -  
R: Around spring time [inaudible]  
FR: Oh, around [month1]?  
R: Let's just pretend it was [month2] or [month1] [laughter]
- 5 – FR: Now, did you get a raise in 2009?  
R: [inaudible] I don't, I don't even look at that stuff, you know [inaudible]  
FR: Ok, ok, so you had a change in pay rate in 2009 ... any idea what month in 2009 you got the raise?  
R: Uh, no idea...

In addition to the single instance of a “textbook” EHC-style response by an FR, in a small number of other cases (n = 4; 25 percent of all assistance attempts) FRs employed a weaker form of memory anchoring to assist recall, by calling the respondent’s attention to standard, calendar-related phenomena, such as well-known holidays, or seasons. This is acceptable practice – such anchors can be helpful – and the 2010 training tried to make that clear, but it is not an EHC “best practice.” Because these phenomena have no obvious salience to the respondent, the intent in EHC interviews is that they should be used only as a fallback option, when better strategies – i.e., known landmarks – are not available. In fact, closer inspection suggests that using generic

calendar phenomena was the FR's best available option in two of the four cases where they used this approach. In one the recall difficulty involved the date of the landmark event itself; there was almost no known, respondent-specific information available to the FR at that point in the interview<sup>17</sup>. In the other case where the generic calendar cue was probably the best available option, the recall difficulty also arose almost immediately. As his second landmark, the respondent reported a domain-relevant event – enrollment in an on-line class – and, rather than focusing on it as a landmark, the FR instead jumped immediately to that domain to follow up. The respondent couldn't immediately recall the start of his enrollment, but thought it might have been in the spring, rendering his one true landmark event, in January, of little use.

In the other two cases where FRs used generic calendar phenomena as recall cues there were clearly better options. In one, although there were no landmarks, there were three different domain events to call upon, any of which might have been more useful than having the respondent try to recall the weather (this case is the example cited for code 4 in Table 7.b.1). In the other case, the situation practically shouted out for the use of the respondent's one landmark event. At the end of the FR's weak probing, with seasonal cues, the start date for a spell of Food Stamps receipt was left as an indeterminate consecutive pair of months – “[month1], [month2]” – the latter of which was the month of the respondent's landmark event.

About two-thirds of FRs' assistance attempts comprised a miscellaneous mix of “other” approaches, the main unifying feature of which was the absence of any effective effort to offer the respondent a useful memory cue. The modal approach in these cases was to probe for a spell start or end date with a direct question on the order of: “What month was that?” (see, e.g., the code 5 example in Table 7.b.3), with no attempt to connect memory for known events in the respondent's life – or even generic calendar phenomena – and the difficult-to-recall event.

### 7.c. The impact of FR assistance on resolving respondent recall problems

For each observed instance of recall difficulty, I also coded the outcome of the interaction (if any) between the FR and the respondent – specifically, whether the difficulty appeared to have been resolved. Obviously, it is impossible to tell from the respondent's verbal behavior alone whether the apparent quandary was *actually* resolved into an accurate report, so my coding focused on evidence of the respondent's apparent confidence in that report. Once again, the coding scheme I devised used a simple three-point scale, with definitely confident and definitely not confident at the end-points, and a middle “maybe/uncertain” category in between. Table 7.c.1 (next page) shows the results.

---

<sup>17</sup> Arguably, an event whose date can't be readily recalled is not much of a landmark, in the generally accepted sense of the term. In this case, however, the event was clearly an important one to the respondent – he became a citizen, a fact that he mentioned immediately after the FR's landmark introduction. Moreover, the FR was very quick to jump in with assistance when the respondent stumbled briefly regarding the date. The interaction proceeded as follows:

FR: [landmark events introduction]

R: Uhh- yes I become a citizen.

FR: You became a citizen last year, ye-yey okay what month was that?

R: Ahh-

FR: Well, was it after New Year's, was it in the winter?

R: Ooh, [month].

---

**Table 7.c.1: Was the Recall Problem Resolved?  
That is, Did the Respondent Settle Confidently on a Response?**

	<u>N</u>	<u>%</u>
1 – yes	9	38
2 – maybe; uncertain	7	29
3 – no	8	33
<b>TOTAL:</b>	<b>24</b>	<b>100%</b>

Coding Examples:

- 1 – FR: Ok, now Food Stamps. Did you receive any Food Stamps in 2009?  
 R: Yes  
 FR: Ok. And what months were those?  
 R: Oh man. [inaudible] so, I have to renew [month1], ok, so, six month before that, so, I think [inaudible] So I want to say, [month2] to [month3], let me see, [month2] to [month3], ok, I got them, I had to renew [inaudible] [month1], expired  
 FR: Ok, so. You got them [month2] through [month3]  
 R: [month2] through [month3] 2009  
 FR: Mhm  
 R: I [inaudible]  
 FR: Were there any months, ah, any weeks in between that you didn't get them?  
 R: Yeah [inaudible] I think the whole month of [month3] I didn't have them.  
 FR: Ok, so, that [month2] through  
 R: [month4] ... 'Cause I know, 'cause 6 months you have to renew
- 2 – FR: You don't remember what month it [period away from job without pay] was?  
 R: Some time maybe in [month].  
 FR: In [month]-ish. Toward the s-beginning of [month], toward the end of [month]?  
 R: Maybe around the middle  
 FR: Ok.
- 3 – FR: Ok, so in [month] last year you probably got overtime is that true do you think?  
 R: I don- I don't remember  
 FR: Ok, well then we'll just put don't know, how's that?  
 R: Yeah, because \*[inaudible]\*  
 FR: \*Because\* we don't really know.  
 R: Yeah...
- 

Note that the number of definitely or possibly resolved recall difficulties – 16 – exactly matches the number of recall difficulties that FRs noticed and took some action to address (see section 7.b, and Table 7.b.1). Is there a connection? Did the fact that FRs offered assistance yield these (presumably) more positive outcomes? A cursory review of the results shown in Table 7.c.2 (next page) suggests that the answer is no. The proportion of definitely or possibly resolved recall difficulties in cases where the FR offered some assistance (69%) does not differ from the proportion successfully resolved in cases where the FR did not offer assistance (63%;  $t=0.29$ , 22 df, n.s.). Based on these results, it seems reasonable to conclude that most recall difficulties are transitory. Table 7.c.2 suggests that, when such a problem arises, there is a good chance that the

**Table 7.c.2: Relationship Between FRs' Assistance Attempts and the Apparent Resolution of Respondents' Recall Difficulties**

		Did the FR Offer Any Assistance?			
		Yes		No	
		N	%	N	%
Was the Recall Difficulty Resolved?	yes/maybe:	11	69	5	63
	no:	5	31	3	38
	TOTAL:	16	100%	8	100%

Statistical analysis summary (significant differences):  
[none]

respondent will work through it on his or her own, without any assistance from the FR, and FR assistance adds little if any value to the process.

A closer look at the data, however, reveals a very different picture – namely, that FR assistance does matter, if it is the right type of assistance. As shown in Table 7.c.3, all of the cases in which the FR addressed the respondent's recall problem with some form of memory anchoring strategy (codes 1 and 4 in Table 7.b.3) eventually saw the recall problem either definitely or possibly resolved. The comparable outcome of resolved recall problems in cases in which the FR employed some other type of assistance not tied to memory processes (code 5) was only about half (55%). Despite the very small numbers of cases, a simple t-test analysis finds this to be a statistically significant difference.

**Table 7.c.3: Relationship Between Type of FR Assistance and the Apparent Resolution of Respondents' Recall Difficulties**

		Type of FR Assistance			
		Memory Anchor (code 1 or 4)		Not Tied to Memory Processes (code 5)	
		N	%	N	%
Was the Recall Difficulty Resolved?	yes/maybe:	5	100 <sup>a</sup>	6	55 <sup>a</sup>
	no:	0	0	5	45
	TOTAL:	5	100%	11	100%

Statistical analysis summary (significant differences):  
<sup>a</sup> t=2.89, 10 df (adjusted for unequal variances), p<.05

7.d. Relationship between FR characteristics (certification test scores and experience/tenure) and whether the FR noticed respondent recall problems

Section 4 of this report examines FR experience/tenure and certification test performance (and their inter-relationships), and section 6 looks at the relationship of those characteristics to the performance of basic EHC interviewing tasks (primarily, eliciting landmark events). The next two results sections investigate whether FRs' prior experience and their success in training (as measured by their certification test scores), are predictive of their taking positive and productive action when presented with respondent recall difficulties. In order to take effective action, of course, a necessary first step is that FRs must notice those difficulties.

Table 7.d.1 presents the mean certification test scores for FRs whose behavior in the recorded interviews suggests that they did or did not notice the respondent's recall difficulty. An effective EHC training program should increase FRs' sensitivity to respondent recall problems, which we would expect to be reflected in higher scores on the certification test among FRs who noticed those problems compared to those who did not. Statistical analysis, however, finds no such differences, even for the sub-scale of test items focused on EHC-specific knowledge. Thus, to the extent that test performance indicates the success of training, there is no indication in these results that the 2010 training served to increase the likelihood that FRs would notice when respondents were experiencing recall difficulty.

---

**Table 7.d.1: Relationship Between Noticing the Respondent's Recall Problem and Certification Test Score**

<u>Mean cert test score for ...</u>	<u>Did the FR Notice the Recall Problem?</u>	
	<u>Yes (N=16)</u>	<u>No (N=6)</u>
... SIPP sub-scale (25 items)	21.3	20.7
... EHC sub-scale (17 items)	14.1	13.8
... Cert Test TOTAL (42 items)	35.4	34.5

Statistical analysis summary (significant differences):  
[none]

---

Table 7.d.2 (next page) examines the relationship between an FR noticing that a respondent was having difficulty recalling the date of some event and his or her prior interviewing experience. These results suggest some real differences among the three experience/tenure groups in whether or not they noticed those difficulties. Somewhat surprisingly, the SIPP-experienced FRs stand out as the least attuned to respondent recall problems. The rate at which they gave evidence of noticing those problems (40%) was significantly lower than the rate for the new hire group (88%), and also significantly lower than the rate for the non-SIPP and new hire groups combined (79%). Standard survey interviewing procedures, such as those adhered to historically by SIPP, train interviewers to be script-readers, primarily. EHC-style interviewing, with its reliance on

**Table 7.d.2: Relationship Between Noticing Respondents' Recall Problems and FR Experience/Tenure**

Did the FR Notice the Recall Problem?	Experience/Tenure					
	SIPP-Experienced		Census (non-SIPP) Experienced		New Hire	
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
Yes	2	40 <sup>a,b</sup>	8	73	7	88 <sup>a</sup>
No	3	60	3	27	1	13
TOTAL:	5	100%	11	100%	8	100%

Did the FR Notice the Recall Problem?	Non-SIPP Experienced and New Hire FRs (combined)	
	<u>N</u>	<u>%</u>
Yes	15	79 <sup>b</sup>
No	4	21
TOTAL:	19	100%

Statistical analysis summary (significant differences):

<sup>a</sup> t=1.92, 11 df, p<.10

<sup>b</sup> t=1.74, 22 df, p<.10

interviewers' on-the-spot ability to recognize and solve problems, represents an important departure from that tradition. EHC interviewing not only permits "off-script" behaviors – it requires them in order to succeed. One possibility is that the performance of the SIPP-experienced FR group suffered because they had to "unlearn" long-ingrained principles, and were not completely successful in doing so in the 2010 test. However, this factor should apply to the non-SIPP Census-experienced FR group as well; the fact that it does not is a clear weakness of an "unlearning" hypothesis.

7.e. Relationship between FR characteristics (certification test scores and experience/tenure) and whether the FR took effective action to assist respondent recall problems

Only after noticing a problem can an FR initiate the second essential step in resolving a respondent's recall difficulty by taking action to assist. The results summarized in section 7.c suggest that some assistance strategies – namely, those which exploit known memory processes, and which comprise the core of the EHC approach – are more effective than others. This final results section looks at FR characteristics as possible predictors of the effectiveness of their responses to respondent recall problems. The very small number of cases – the 16 instances in which FRs made any attempt to assist a respondent having difficulty recalling the date of an

event – hampers the analysis considerably; the problem is only exacerbated in the further subdivision of those few cases into various comparison groups. Nevertheless, while certainly not definitive, the results are intriguing.

Table 7.e.1 examines the relationship between the type of assistance FRs offered and their end-of-training certification test performance. The results offer no evidence that test performance was related to use of the more effective, “memory anchor” assistance types. The absence of any predictive potential in the EHC sub-scale in particular is somewhat surprising, although these results parallel those in section 7.d about noticing respondent recall problems – both phenomena appear to have been unrelated to FRs’ scores on the EHC-related test items. Therefore, unless the test was seriously flawed, the suggestion in these results is that the training program was not effective in teaching FRs how to respond to a respondent’s recall difficulty. It is worth noting again, however, that the numbers are very small and warrant no definitive conclusions.

---

**Table 7.e.1: Relationship Between Type of FR Assistance and Certification Test Score**

<u>Mean cert test score for ...</u>	<u>Type of FR Assistance</u>	
	<u>Memory Anchor (code 1 or 4) (N=5)</u>	<u>Not Tied to Memory Processes (code 5) (N=10)</u>
... SIPP sub-scale (25 items)	21.8	20.8
... EHC sub-scale (17 items)	13.4	14.4
... Cert Test TOTAL (42 items)	35.2	35.2

Statistical analysis summary (significant differences):  
[none]

---

Table 7.e.2 (next page) examines the relationship between the type of assistance FRs offered and their prior experience (or not) as Census Bureau interviewers. Here again, comparisons among the three experience/tenure groups are rendered very difficult by the almost vanishingly small numbers. In spite of this problem, however, statistical analysis suggests a pattern – namely, the notably poorer performance of the newly hired FRs compared to the two experienced groups. In no case did a new hire FR who acted upon a perceived respondent problem use an assistance strategy with good prospects for success; in combination, the two experienced groups used these strategies half the time. These results do not exactly contradict the experience/tenure results for “noticing” respondent problems, where SIPP-experienced FRs appeared to underperform relative to the rest, but they are certainly somewhat at odds with those results. Perhaps the “noticing” results are an anomaly, since the current findings are much more in line with the earlier results (see section 6) showing less proficiency among the new hires regarding some of the basic tasks of EHC interviewing (primarily, those having to do with eliciting landmark event reports). Or perhaps the same factors that elicit higher rates of “noticing” recall problems – related, in some way, to the newness of the situation – are less likely to yield effective, on-the-spot solutions.

**Table 7.e.2: Relationship Between Type of FR Assistance and FR Experience/Tenure**

<u>Type of Assistance</u>	<u>Experience/Tenure</u>					
	<u>SIPP-</u>		<u>Census</u>		<u>New Hire</u>	
	<u>Experienced</u>		<u>Experienced</u>			
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
Memory anchor (code 1 or 4)	1	50	4	50 <sup>a</sup>	0	0 <sup>a,b</sup>
<u>Not tied to memory processes (code 5)</u>	<u>1</u>	<u>50</u>	<u>4</u>	<u>50</u>	<u>6</u>	<u>100</u>
TOTAL:	2	100%	8	100%	6	100%

<u>Type of Assistance</u>	<u>SIPP and Census (non-SIPP)</u>	
	<u>Experienced FRs (combined)</u>	
	<u>N</u>	<u>%</u>
Memory anchor (code 1 or 4)	5	50 <sup>b</sup>
<u>Not tied to memory processes (code 5)</u>	<u>5</u>	<u>50</u>
TOTAL:	10	100%

Statistical analysis summary (significant differences):

<sup>a</sup> t=2.65, 7 df (adjusted for unequal variances), p<.05

<sup>b</sup> t=3.00, 9 df (adjusted for unequal variances), p<.05

## 8. Conclusions

This final section uses the research results to draw conclusions about the 2010 SIPP-EHC instrument and procedures, and the training program that attempted to produce FRs who were adept at administering them. The conclusions must be tentative because they are based on research with numerous limitations, chief among which are the following: the small number of interviews selected for recording; the absence of a rigorous, statistically defensible method for selecting those cases; the similar absence of any sort of sample of FRs; the use of a highly subjective coding system with only a single coder; and the possibility (suggested by some results) that recording the interviews may have affected how they were conducted. All of these limitations have been at least touched upon earlier in this report, but they merit mention again in the context of drawing conclusions from the research. It seems reasonable to assume that whatever biases the study's limitations impart to the results are likely to be in the direction of having produced more positive, "by-the-book" interviewing behaviors than would be found in a true sample of cases unaffected by the presence of the recorder. If this is the case, then the study might well miss some true problems, due to the masking effects of only examining the experiences and behaviors of highly motivated FRs operating under optimal conditions. On the other hand, that same logic would suggest that any problems that do work their way to the surface under those conditions are very likely to be real, and their magnitudes very likely to be underestimated. Assumptions are not facts, however, and the true impact of the study's many

limitations remain unknown. In offering a set of conclusions and recommendations, as I do below, I simply urge that they be treated with some caution.

(1) Drop the calendar aid as an active tool for respondents' use. It would no doubt be useful to respondents to have available, and be able to refer to, a standard 12-month calendar. However, the procedures in the 2010 test, which called for FRs to attempt to engage them actively in making entries on the calendar, were almost a complete failure. The calendar aid recommendation would undoubtedly be quite different – something akin to “stronger training needed, and more training emphasis” – had the interviews offered more evidence of a demand for strong tools to assist respondent recall. Since they did not – since respondent recall issues appeared only infrequently – there seems to be little reason to retain the procedure. As noted earlier, the calendar aid was not included in the 2011 field test, and at present there are no plans to reinstate it.

(2) Landmark events do not seem to be essential to the SIPP-EHC interview. First, the 2010 results suggest that about a third of the events that respondents report as landmarks are likely to duplicate material soon to surface in the main substantive domains of the EHC interview. In an interview that draws some criticism for excessive length, it is hard to justify spending time covering much the same material in two different places. Second, however, even when the landmarks refer to events that do not overlap with the substantive domains, the issue remains that there does not seem to be much call for their use. And dropping landmarks does not leave the interview empty of potential memory aids. As noted earlier, landmarks have no unique status in that regard – any event recalled by the respondent with confidence can serve the same purpose.

For largely the same reasons, managers of the University of Michigan's Panel Study of Income Dynamics (PSID) recently decided to drop landmarks from its EHC interview (McGonagle, 2012), so there is precedent for this action. Another factor that renders SIPP-EHC well suited not to bother with landmarks – perhaps uniquely well suited among EHC interviews – is its short reference period. Other EHC research which has concluded that landmarks are an important feature of EHC methods, and do contribute to response accuracy (e.g., Belli et al., 2001; but see van der Vaart and Glasner, 2007, for an alternate view), has all been conducted in the context of surveys covering a much longer reference period than the single year of the SIPP-EHC interview.

Although they may not be essential to the SIPP-EHC interview, landmark events may still serve a positive, though not easily quantifiable, function. Anecdotal reports from the field suggest that conversations between FRs and respondents about landmark events have rapport-building benefits, which would argue for retaining them in the SIPP-EHC procedures. It is certainly the case that landmarks are often interesting and dramatic things to talk about – perhaps especially in contrast to the dry nature of most of the SIPP interview's content. They also allow respondents to talk freely about themselves in ways that the rest of the interview does not. In the end, whether these “soft” considerations trump the research-based suggestion to eliminate landmarks as a separate category of inquiry in the interview would seem to be a judgment call, beyond the scope of this paper.

(3) If subsequent iterations of the SIPP-EHC interview retain landmarks, the results point to several training needs with regard to how FRs should behave in attempting to elicit them from respondents, as follows:

(a) Negative landmarks: Negative and even tragic events surfaced fairly often in respondents' landmark event reports – between one-quarter and one-third of the time in the recorded interviews. Far too often in these circumstances FRs failed to respond with even the most rudimentary expression of sympathy; Moore (2011) cites several painful examples. The 2010 training did cover what to do when respondents reported very negative events, but for many FRs that training appears to have been ineffective. It needs to be strengthened.

(b) Early-in-the-year landmarks: The training message to FRs that they should take special (but very simple) steps to try to elicit landmarks from the early months of the reference year seems to have largely been ignored. Faced with a situation in which probing for early landmarks would have been appropriate, about 80% of FRs failed to do so. This is somewhat surprising in light of the evidence that the 2010 training was at least momentarily effective in stressing importance of early landmarks – over 80% of responses to the certification test item dealing specifically with the matter (item 24) were correct. “Under fire,” however, FRs mostly failed to produce the desired action.

(c) Landmarks in proxy interviews: For the most part, FRs failed to follow the 2010 procedures which deemed a single attempt to elicit landmark events sufficient for any one respondent, even when that respondent went on to provide proxy data for another household member. In this case, the shortcomings of the 2010 training were evident even before interviewing began – almost three-quarters of the FRs responded incorrectly in the certification test (see item 21) that they should re-ask the landmarks question in a proxy interview situation. This misunderstanding – especially as revealed by the awkward and bizarre nature of many of the proxy landmark requests that FRs made – is troubling. It suggests not just a failure to follow rules, but a much more fundamental failure to really “get” what landmarks are – not just another data item, but rather a potential anchor point *in the head of* the person doing the reporting, to be called upon as needed to assist more accurate recall *from that person's memory*, regardless of whether he or she is the person being reported *about*. FRs need to leave training with the proper understanding of what to do about landmarks in a proxy interview. Successfully imparting that understanding may depend as much or more on better educating FRs about their fundamental nature and purpose as on the rote specifics of what to do about them in proxy interviews.

A quite different conclusion may merit some consideration. The reasons why devoting time to special probing for landmarks may be unwise – limited interview time, duplication with domain reports, minimal demand for their use – certainly argue even more strongly against asking about them twice. But perhaps, rather than not asking for landmarks in proxy interviews, they should only be asked about in those more challenging interview circumstances, where a respondent's need for assistance is presumably greatest. If so, it might be desirable to try to focus such “proxy” landmarks on significant events involving the proxied-for person. Of course, to the extent that proxy reporting difficulties derive from the absence of information, as opposed to memory retrieval issues, then applying resources to memory processes is unlikely to be fruitful. In other words, landmarks may not be of much benefit when, for example, one temporary

roommate is proxying for another; they may be, however, in the case of a parent reporting for a child. Obviously, evaluating this approach is outside the scope of the current paper, and determining its merits would require additional research.

(4) The major “good news” in the analysis of the recorded interviews is that overt evidence of respondent recall problems in the SIPP-EHC interview appeared relatively rarely. It is important to stress again that this is not the same as saying that actual recall problems were rare, or that respondents reported most matters accurately – the research was not designed to address those issues. But the worry going into the test was that the new, much-longer, 12-month recall period, compared to SIPP’s traditional 4-month period, might be very burdensome to respondents, and might result in major decreases in quality. On the surface, at least, in the observable behavior of the respondents, those fears were not borne out<sup>18</sup>.

In retrospect, perhaps this should not have been surprising. As noted before, the history of the development of EHC methods seems to have occurred exclusively in the context of surveys with much longer reference periods than a year. Although in relative terms a year is a notable step up from the current 4-month SIPP reference period, in absolute terms it may not be a particularly challenging length of time over which to recall life events of the type that are of interest to SIPP. On the surface, the interactions captured in the 2010 recordings suggest that this is the case. Another factor to consider is that in some ways the new reference period, even though longer than the old one, may be easier for respondents, and may actually support better recall. Unlike random sets of four consecutive months, a calendar year makes simple intuitive sense as a period of time to think back about, and year ends and beginnings themselves make natural and easily understandable anchor points.

(5) Infrequent evidence of respondent recall problems is not the same as no evidence of recall problems. Although it was surprisingly rare, such evidence did occasionally appear in the 2010 test. The major strength of the EHC method is that it provides interviewers with a strong tool to address those problems when they appear, and to do so in a way that assists memory without biasing the respondent’s report. In the first stage of that process – recognizing a problem signal when it appeared – FRs in the 2010 test could have performed better than they did. And in the second stage – taking effective action – there was even greater room for improvement. In order to fully exploit the strengths of the EHC method, future iterations of the FR training program must improve FRs’ ability to recognize recall problems when they occur, and, when they do perceive a problem, must better equip them to call on effective strategies to assist recall. These conclusions hold for all FR experience levels, but probably especially for those with no prior interviewing experience.

The fact that FRs failed to do what we wanted them to do leads almost tautologically, *ipso facto*, to the conclusion that their training for the 2010 test was inadequate. The present research provides no hard evidence about the nature of that inadequacy, so the thoughts that follow are highly speculative. First, I suspect that some improvements could be made in how we train FRs to do what we want them to do – the content and focus of the EHC-specific training material –

---

<sup>18</sup> And below the surface as well. Gathright and colleagues (SIPP-EHC Data Evaluation Workgroup, 2011), in an examination of 2010 field test estimates and comparable estimates derived from the 2008 panel of the production SIPP survey, find much evidence of very comparable levels of data quality.

mostly because it is hard to imagine that we got it exactly right on our first attempt. But I also suspect that there is a more fundamental, root problem to be addressed. EHC methods require that interviewers pay sufficient attention to a respondent's words and actions to be able to perceive sometimes subtle signs of recall difficulty. And, once they perceive those signs, they must think their way to an effective remedy without the benefit of a script. Compared to traditional interviewing practices – reading questions and recording answers – these are difficult skills to master. Both require intensive and repeated practice, and both require that some minimum cognitive capacity be available during interviewing to devote to them. Abundant anecdotal evidence suggests that the 2010 SIPP-EHC instrument was far from smooth and polished, with many unclear or otherwise poorly designed individual screens, as well as some interview sections with major usability issues. It is almost certain that these instrument flaws detracted from both EHC skill-building practice time in training and from FRs' cognitive capacity to deal effectively with respondent recall problems as they arose in the field. More attention to making the operation of the instrument easy and obvious would mean more time in training to devote to matters that don't have to do with the scripted interview, and would also free up cognitive capacity for FRs to be more alert to problems in the field and to come up with quick and effective solutions.

There is evidence that instrument refinements following the 2010 test, along with increases in FR comfort and familiarity with the instrument which naturally attend additional experience with it, are succeeding in making the instrument more “easy and obvious” for FRs to use. Specifically, Walsh and Fields (2012) report a substantial reduction in average interview time in the 2011 SIPP-EHC field test compared to the 2010 experience. In addition, Walsh (2012) finds evidence that the 2010 experience of many of the 2011 FRs benefitted their 2011 interviewing performance. These trends, along with improvements to the training itself (FLD training specialists and others have expressed confidence that the training program was improved for 2011), augur well for the future. Unfortunately, the 2011 field test included no interview recording, so a direct assessment of the effects of these factors on cornerstone EHC interviewing practices awaits future research.

(6) The end-of-training certification test, although an imperfect predictor of subsequent EHC interviewing performance, was still a surprisingly good one, especially given that both the training and the test which attempted to assess its success were inaugural efforts. Performance on the subset of test items dealing with EHC-specific (as opposed to SIPP-specific) matters was consistently and positively associated with a number of specific, desired EHC interviewing behaviors. This was not the case, however, for the more complex and demanding behaviors that form the core of successful EHC-style interviewing. That is, the test was unrelated to whether or not the FR noticed that a respondent was experiencing recall difficulty, and also unrelated to whether the FR took effective, EHC-style action. An end-of-training test that was predictive of these skills would allow follow-up training to be focused on FRs most in need of it. Again, post-2010 modifications to the test may be moving it in the right direction. Walsh (2012) finds evidence in the 2011 field test results of better predictive power for the certification test for some FR performance indicators, although it is unclear whether that improvement extends to the sorts of EHC behaviors examined in the present research.

(7) Interview recording is a valuable addition to field test design, and should be included in all future tests of SIPP-EHC procedures. The data that recordings generate are laborious and time-consuming to produce. Nevertheless, in many ways they are unsurpassed in their ability to reveal how refinements in instrument design and training, and FRs' steady accretion of experience with the instrument and with EHC procedures in general, affect performance in the field.

## References

- Armor, D. and S. Taylor (2002), "When Predictions Fail: The Dilemma of Unrealistic Optimism," in Gilovich, T., D. Griffin, and D. Kahneman (eds.) (2002), Heuristics and Biases: The Psychology of Intuitive Judgment, Cambridge, UK: 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, 383-406.
- Belli, R., W. Shay, and F. Stafford (2001), "Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods," *Public Opinion Quarterly*, 65, 45-74.
- Belli, R. and F. Stafford (2008), "Future Directions in Calendar and Time Diary Methods," in Belli, R., F. Stafford, and D. Alwin (eds.), Calendar and Time Diary Methods in Life Course Research. Thousand Oaks, CA: Sage Publications, 2008.
- Edwards, A. (2012), "An Evaluation of Poverty in the SIPP-EHC," paper presented at the Annual Meeting of the Population Association of America, San Francisco, CA, May 3-5, 2012.
- Ellis, R. and Chenevert, R., (2012), "'I Don't Remember': Effects of Recall Period on Reported Job and Program Participation Duration," paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC, January 10-12, 2012.
- Fields, J. and M. Callegaro (2007), "Background and Planning for Incorporating an Event History Calendar into the Re-Engineered SIPP," paper prepared for presentation at the Federal Committee on Statistical Methodology Statistical Policy Seminar, Washington, DC, November 2007.
- Fields, J. and R. Walsh (in preparation), "2010 SIPP-EHC Evaluation – Design, Implementation, Analysis, and Summary," unpublished U.S. Census Bureau paper.

- Hillmer, B., L. Monte, and J. Fields (2011), "Landmark Event Reporting in the 2010 Survey of Income and Program Participation Event History Calendar," poster presentation at the Annual Meeting of the Population Association of America, Washington, DC, March 31 – April 2, 2011.
- Marlay, M. and P. Mateyka (2012), "A Preliminary Evaluation of the Residence History Data in the 2010 and 2011 Field Tests of the Reengineered Survey of Income and Program Participation," paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC, January 10-12, 2012.
- Matlin, M., and D. Stang (1978), The Pollyanna Principle: Selectivity in Language, Memory, and Thought, Cambridge, MA: Schenkman.
- McGonagle, K. (2012), personal communication, June 5, 2012.
- Moore, J. (2011a), "2010 SIPP-EHC Field Test Transcript Review: Part 1: Notes Regarding Instrument Design Issues," unpublished U.S. Census Bureau report, September 14, 2011.
- Moore, J. (2011b), "2010 SIPP-EHC Field Test Transcript Review: Notes Regarding FR Training Issues," unpublished U.S. Census Bureau report, November 9, 2011.
- Moore, J., J. Fields, G. Benedetto, M. Stinson, A. Chan, and J. Maples (2009), "The 2008 Survey of Income and Program Participation Event History Calendar Field Test: Study Design and Initial Results," paper prepared for presentation at the annual meetings of the American Association for Public Opinion Research, Hollywood, FL, May 14-17, 2009.
- National Research Council (2009), "Reengineering the Survey of Income and Program Participation," Panel on the Census Bureau's Reengineered Survey of Income and Program Participation, Citro, C. and J. Scholz (eds.), Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC.
- Pascale, J., J. Moore, and A. Chan (2010), "Re-Engineered Survey of Income and Program Participation Event History Calendar 2010 Field Test Transcription Training Manual," unpublished US Census Bureau (Statistical Research Division) document, January 20, 2010.
- SIPP-EHC Data Evaluation Workgroup (2011), "An Initial Evaluation of the 2010 Field Test of Re-Engineered SIPP," U.S. Census Bureau, SEHSD Working Paper 2011-9, March 2011.
- Stinson, M., G. Gathright, and J. Skog (2012), "Comparing Job Characteristics from the 2010 SIPP-EHC Field Test to the Census Bureau Business Register," paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC, January 10-12, 2012 (paper dated March 20, 2012).

- U.S. Census Bureau (2009), “Re-Engineered Survey of Income and Program Participation (RE-SIPP) Field Representative’s Interviewing Manual – 2010 Field Test,” unpublished U.S. Census Bureau document, October 2009.
- van der Vaart, W., and T. Glasner (2007), “The Use of Landmark Events in EHC-Interviews to Enhance Recall Accuracy,” paper presented at the U.S. Census Bureau/Panel Study of Income Dynamics Conference on the Use of Event History Calendar Methods in Panel Surveys, Washington, DC, December 5-6, 2007.
- Walsh, R. (2012), “Interviewer Effect on Conversational Interviews: Results from the 2010 & 2011 SIPP-EHC,” paper presented at the Annual Conference of the Population Association of America, San Francisco, CA, May 3-5, 2012.
- Walsh, R. and J. Fields (2012), “Reducing Respondent Burden: Evaluating the Progress of the SIPP-EHC,” paper presented at the Federal Committee on Statistical Methodology (FCSM) Research Conference, Washington, DC, January 10-12, 2012.

## ATTACHMENT A

### 2010 SIPP-EHC Field Test – FR End-of-Training “Certification Test”

#### Notes:

- (1) The test booklet is labeled as the “2009 ... End of Training Concepts Test.” Training was administered late in calendar year 2009 for the 2010 field test, for which interviewing began in January 2010.
- (2) The appendix copy of the test booklet also shows the percent distribution of responses to the test items among the 269 FRs who completed the test. Correct answers are indicated by a percentage figure highlighted in bold.

**2009 SURVEY OF INCOME AND PROGRAM  
PARTICIPATION - EVENT HISTORY CALENDAR****End of Training Concepts Test  
Frequency of Responses n=269**

FR Name:

FR Code:

Trainer Name:

Today's Date:

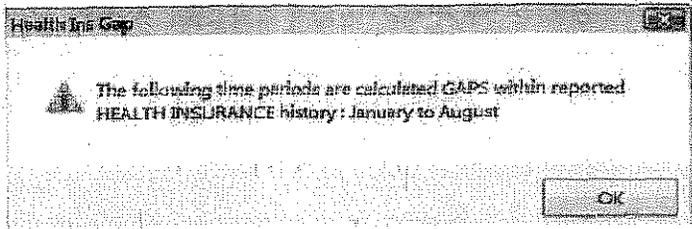
Regional Office:

**PART 1. SIPP CONCEPTS** *Select the best response for each question*

1) Which law establishes the Census Bureau's requirements for maintaining data confidentiality? <i>(Circle only one)</i>	a) Title 5, The Privacy Act 2.2% b) Title 13, of the United States Code 97.4% c) Title 26, of the United States Code 0.0% <i>blank</i> 0.4%
2) Which of the following is <b>not</b> a criteria for determining if living quarters meet the housing unit definition for SIPP? <i>(Circle only one)</i>	a) The housing unit has access through an outside entrance 3.4% b) The housing unit has access through a common hallway 5.6% c) The housing unit has a unique address 80.3% d) Occupants live separately from other households on the property 9.7% <i>blank</i> 1.1%
3) Which type of Group Quarters are eligible for SIPP interviewing? <i>(Circle only one)</i>	a) Institutional 7.4% b) Non-institutional 87.7% c) Military Barracks 4.1% <i>blank</i> 0.7%
4) Which of the following is the best response when a respondent asks "Why was I selected for the survey?" <i>(Circle only one)</i>	a) The economic status of this household meets the requirements for the survey 3.7% b) A computer scientifically selects people at random to be in the sample 21.2% c) Your address was selected to be in the sample, not you individually 74.7% d) The supervisor in the office selects which addresses are in sample 0% <i>multiple response</i> 0.4%
5) When should you begin contacting your SIPP cases? <i>(Circle only one)</i>	a) At the beginning of the interview period to determine if there are any problems 98.5% b) In the middle of the interview period once all cases have been downloaded 0.4% c) At the end of the interview period when respondents are more likely to be available 0.4% <i>blank</i> 0.7%
6) Who is an eligible respondent for SIPP? <i>(Circle only one)</i>	a) A permanent household member 12 years of age or older 0.4% b) A permanent household member 15 years of age or older 98.5% c) A permanent household member 18 years of age or older 0.7% <i>multiple response</i> 0.4%

<p>7) Who is the reference person for the SIPP interview? (Circle only one)</p>	<p>a) The initial respondent 9.7% b) Any household member 15 years of age or older 6.0% c) The person or persons who own or rent the home 84.0% d) The oldest person in the household 0.0% multiple response 0.4%</p>																																
<p>8) Who is a "proxy" respondent? (Circle only one)</p>	<p>a) A household member who is not available at the time of interview 1.9% b) A person who is responding for a household member who is not available 94.8% c) A household member under 15 years old whose parents respond to the interview for him. 2.6% blank 0.4% multiple response 0.4%</p>																																
<p>9) Which of the items listed is asked for all household members? (Circle only one)</p>	<p>a) Age 83.6% b) Marital status 7.1% c) Educational attainment 6.3% d) Service in the U.S. Armed Forces 0.0% blank 0.4% multiple response 2.6%</p>																																
<p>10) Which of the terms listed refers to the areas of the respondent's life that are covered in the Event History Calendar? (Circle only one)</p>	<p>a) Domains 69.9% b) Spells 25.7% c) Scenarios 1.9% blank 1.5% multiple response 1.1%</p>																																
<p>11) For each of the items below, determine whether or not the item is considered income. (Mark "Yes" if the item is considered income, mark "No" if the item is not considered income.)</p> <table border="0" style="width: 100%;"> <thead> <tr> <th style="text-align: left;">Yes</th> <th style="text-align: left;">No</th> <th style="text-align: left;">blank</th> <th></th> </tr> </thead> <tbody> <tr> <td>a. 46.1%</td> <td>52.4%</td> <td>1.5%</td> <td>Room and board paid each month by the respondent's adult son living in a basement bedroom</td> </tr> <tr> <td>b. 63.9%</td> <td>34.6%</td> <td>1.5%</td> <td>Income from a rental property when the expenses exceed the amount of rent received</td> </tr> <tr> <td>c. 84.0%</td> <td>14.5%</td> <td>1.5%</td> <td>Interest on a savings account you own jointly with your 16 year old son</td> </tr> <tr> <td>d. 24.9%</td> <td>72.5%</td> <td>2.6%</td> <td>\$2,000 you receive each year on your birthday from your grandmother in a retirement home</td> </tr> <tr> <td>e. 44.2%</td> <td>52.8%</td> <td>3.0%</td> <td>Money you receive under the G.I. bill to pay for college tuition</td> </tr> <tr> <td>f. 53.9%</td> <td>42.8%</td> <td>3.4%</td> <td>\$25,000 inheritance received upon the death of your great uncle</td> </tr> <tr> <td>g. 10.4%</td> <td>87.0%</td> <td>2.6%</td> <td>A teenage son's weekly allowance</td> </tr> </tbody> </table>		Yes	No	blank		a. 46.1%	52.4%	1.5%	Room and board paid each month by the respondent's adult son living in a basement bedroom	b. 63.9%	34.6%	1.5%	Income from a rental property when the expenses exceed the amount of rent received	c. 84.0%	14.5%	1.5%	Interest on a savings account you own jointly with your 16 year old son	d. 24.9%	72.5%	2.6%	\$2,000 you receive each year on your birthday from your grandmother in a retirement home	e. 44.2%	52.8%	3.0%	Money you receive under the G.I. bill to pay for college tuition	f. 53.9%	42.8%	3.4%	\$25,000 inheritance received upon the death of your great uncle	g. 10.4%	87.0%	2.6%	A teenage son's weekly allowance
Yes	No	blank																															
a. 46.1%	52.4%	1.5%	Room and board paid each month by the respondent's adult son living in a basement bedroom																														
b. 63.9%	34.6%	1.5%	Income from a rental property when the expenses exceed the amount of rent received																														
c. 84.0%	14.5%	1.5%	Interest on a savings account you own jointly with your 16 year old son																														
d. 24.9%	72.5%	2.6%	\$2,000 you receive each year on your birthday from your grandmother in a retirement home																														
e. 44.2%	52.8%	3.0%	Money you receive under the G.I. bill to pay for college tuition																														
f. 53.9%	42.8%	3.4%	\$25,000 inheritance received upon the death of your great uncle																														
g. 10.4%	87.0%	2.6%	A teenage son's weekly allowance																														
<p>12) In what order are you required to conduct the subject domains? (Circle only one)</p>	<p>a) You must go through the domains in the order they occur in the instrument 2.6% b) You may go through the domains in any order 96.3% blank 1.1%</p>																																

13)



The pop-up screen shown indicates a gap in data collected for the Health Insurance domain. What should you do? (Circle only one)

- a) Click "OK" and continue with the interview 10.0%
- b) Be aware that there must not be gaps for the domain upon completing the EHC 45.7%
- c) Inquire about gaps immediately each time you see the message 42.4%
- blank 0.7% multiple response 1.1%

14) Match the Case Management function key (in the left column) to the description (in the right column) by entering the corresponding letter in the answer space. (Enter one letter for each function key)

percent correct

- |       |       |     |   |
|-------|-------|-----|---|
| 95.2% | _____ | F1  | A. Begins an interview for the case that is currently highlighted on your case list |
| 96.3% | _____ | F2  | B. Allows you to add notes and view existing notes for a highlighted case.          |
| 94.4% | _____ | F7  | C. Exits Case Management  |
| 97.4% | _____ | F10 | D. Access a general help screen   |

15) For each of the following scenarios, classify the noninterview as a Type A Noninterview, Type B Noninterview, Type C Noninterview or Type Z Noninterview (Mark (X) the appropriate box for each scenario).

Noninterview Type:

- |    | A     | B     | C     | Z     | blank |  |
|----|-------|-------|-------|-------|-------|--|
| a. | 77.3% | 17.8% | 0.7%  | 1.5%  | 2.6%  | The unit is occupied but an interview was not possible because the respondents were on vacation. |
| b. | 10.0% | 6.0%  | 3.4%  | 76.6% | 4.1%  | One of three roommates in an apartment refuses to be interviewed.                                |
| c. | 1.9%  | 34.9% | 59.1% | 0.7%  | 3.4%  | A vacant house that is permanently being used as storage for John Smith's printing business.     |
| d. | 8.6%  | 46.1% | 22.3% | 15.6% | 7.4%  | A house that is occupied entirely by armed forces members  |
| e. | 2.2%  | 13.4% | 76.2% | 4.1%  | 4.1%  | A house that was converted to a permanent business   |

## PART 2. LANDMARK SCENARIO QUESTIONS

*For the Landmark scenarios in questions 16 through 21, chose the action that best describes how you should respond to the Landmark question in that particular situation*

**Landmark question:** Today it is January of 2010. In this section of the interview we are interested in events that have occurred during the last calendar year, that is from January thru December of last year (2009). We are especially interested in where and with whom you have lived, your employment and unemployment history, and your income from work and from programs. First, though, are there any events during the last calendar year, that is from January thru December of last year (2009), that stand out in your mind that you are able to date exactly or approximately? These might be family, financial, job, social or health related events.

**16) Landmark Scenario 1:** In response to the landmark question, the respondent replies: "I can't think of anything. There wasn't anything important, I don't think." (Circle the best action.)

<p>a) <i>Using your judgment, and whatever you already know about the respondent, supply two or three landmarks yourself:</i> “Well, you said your birthday is in February, so let’s use that. And there’s always the 4<sup>th</sup> of July. And I assume your kids went back to school in September? We’ll use that, too.” 12.6%</p>	<p>b) <i>Probe gently:</i> “Are you sure? It doesn’t have to be anything real important, just something memorable to you. Like a trip somewhere, or maybe somebody visited you [or work-related, or health-related, or family-related]...” <i>If gentle probing fails, move on.</i> 82.5%</p>
<p>c) <i>It is very unlikely that someone would have no important life events over an entire year. Probe with gradually increasing pressure until he/she produces at least one landmark (or, preferably, two or three). If probing fails, use your judgment, and whatever you already know about the respondent, and supply two or three landmarks yourself</i> 0.7%</p>	<p>d) <i>Back away politely and move on:</i> “Never mind; don’t worry about it; it’s not important. Let’s move on.” 1.1%</p> <p><i>blank 1.5% multiple response 1.5%</i></p>
<p>17) <b>Landmark Scenario 2:</b> <i>In response to the “Landmarks” question, the respondent replies “That’s a strange question. Why do you need to know that?” (Circle the best action.)</i></p>	
<p>a) <i>Explain:</i> “This survey is about the economic situation of people in the US. We ask these questions so that researchers and policy-makers can develop the best policies to help people in need.” 17.5%</p>	<p>b) <i>Address the respondent’s concerns, explain, and re-ask the question:</i> “We’re not being nosy. And don’t worry – we ask everyone this question. An ‘important event’ is whatever that means to you. So, can you think of any important events in your life in the last year...?” 11.5%</p>
<p>c) <i>Explain:</i> “We do this just to help people get started thinking about the last year, and what happened during the year that might help them remember other things. They’re like anchor points in your memory.” 69.9%</p>	<p>d) <i>Back away politely and move on:</i> “Never mind; don’t worry about it; it’s not important. Let’s move on.” 0.0%</p> <p><i>blank 0.4% multiple response 0.7%</i></p>

18) **Landmark Scenario 3:** *In response to the “Landmarks” question, the respondent replies: “Well, I was in the hospital for a week in February for a ruptured appendix. I won’t forget that for a long time. And another big thing was that my mom, who’s been alone for almost ten years since dad died, got re-married – on Halloween! Oh yeah, and my husband and I finally saw the Grand Canyon this summer. What a beautiful place. That was in July – no, I’m wrong, it was late June we were there.” (Circle the best action.)*

<p>a) <i>Probe for additional events, especially something early in the year:</i> “Good. Can you think of two or three more? Anything else? Anything related to your work, or other members of your family, or other trips you took, or maybe somebody visited you...? Any other events that you remember, especially early in the year, would be useful.” 8.6%</p>	<p>b) <i>Enter the events in the EHC, and insist that the respondent write them down on the calendar aid. Proceed to the next topic only after you’re certain he/she has entered everything on the calendar:</i> “Thank you. So the hospital stay was in February. Write that down on that calendar I gave you, on the ‘Life Events’ line. And write ‘Mom’s wedding’ in October, and ‘Grand Canyon’ in June. Do you have all three of those on the calendar? Good; let’s move on to the next topic...” 6.3%</p>
<p>c) <i>Enter the events in the EHC, and encourage the respondent to write them down on the calendar aid:</i> “That’s great – thanks. So the hospital was in February... Why don’t you write that down on that calendar I gave you? Your mom’s wedding was October. And the Grand Canyon was in June, you decided. Great!” 78.1%</p>	<p>d) <i>Using your best judgment, select the two best landmarks and enter them on the EHC:</i> “OK, let’s use the hospital stay in February and your mother’s wedding in October – those sound like your two most memorable events.” 5.6%</p> <p><i>multiple response</i> 1.5%</p>

**19) Landmark Scenario 4:** *In response to the “Landmarks” question, the respondent replies: “My dog died. That was tough. I had him for 12 years – got him from the pound when I was 16.” (Circle the best action.)*

<p>a) <i>This is not a useful landmark – it’s very negative, and probably very distressing to the respondent, and there’s no information about the month in which it occurred. But, because it’s unlikely that the respondent has anything <u>more</u> memorable than this to report, do not probe for other landmarks. Supply two or three yourself, using what you already know about the respondent and/or commonly-understood ‘generic’ landmarks such as Easter, the 4<sup>th</sup> of July, Christmas, etc.</i> 1.9%</p>	<p>b) <i>Feel free to express sympathy. Probe for the month; enter the event in the EHC, and encourage the respondent to write it down on the calendar aid; probe gently for any other landmarks (especially early in the year, if necessary):</i> “I’m sorry to hear that. When was that? [Respondent provides the month] OK, I’m going to put that in my computer, and why don’t you jot that down on that calendar I gave you. Anything else you can think of that were important events in your life during 2009?” 86.3%</p>
<p>c) <i>This is not a useful landmark – it’s very negative, and probably very distressing to the respondent, and there’s no information about the month in which it occurred. Probe for other landmarks:</i> “Let’s try for something else. Is there anything else you can think of that were important events in your life during 2009?” 3.0%</p>	<p>d) <i>Feel free to express sympathy. Using the information the respondent has provided, probe to determine the date of the landmark:</i> “Oh, that is hard, I know. Do you remember what month you got him from the pound? And was it <i>exactly</i> twelve years that he lived with you?” 8.2%</p> <p><i>multiple response</i> 0.7%</p>

**20) Landmark Scenario 5:** *In response to the “Landmarks” question, the respondent replies: “Let’s see.... Last fall I finally got rid of my car; I was just thinking about this earlier today. It was the day after Thanksgiving. I’ve been thinking for years that I could live without one, and bang! I finally just did it; sold it. No regrets so far. In fact, it’s been great.” (Circle the best action.)*

<p>a) Enter the event in the EHC, encourage the respondent to write it down on the calendar aid, and probe for anything else, especially early in the year: "All right. That calendar I gave you? On the Life Events line why don't you write down 'no car' in November. Was there anything else that you remember? Anything in the winter or spring, especially? About your family, or work, or your health...?" 78.4%</p>	<p>b) Enter the event in the EHC, and encourage the respondent to write it down on the calendar aid. Events like this are often tied to changes in residence – a move to a home close to public transportation, for example – so move directly to the "Residence" domain: "All right. That calendar I gave you? On the Life Events line why don't you write down 'no car' in November. Did you move to this place at about the same time you sold your car?" 14.1%</p>
<p>c) Enter the event in the EHC, and encourage the respondent to write it down on the calendar aid. Events like this are often tied to changes in economic circumstances, such as the loss of a job, so move directly to the "Jobs" domain: "All right. That calendar I gave you? On the Life Events line why don't you write down 'no car' in November. Did you lose a job at about that time, so you could no longer afford to keep a car?" 5.6%</p>	<p>d) Enter the event in the EHC, encourage the respondent to write it down on the calendar aid, and move on: "All right. That calendar I gave you? On the Life Events line why don't you write down 'no car' in November? And that's plenty so let's move on." 0.7%</p> <p>blank 0.4% multiple response 0.7%</p>

**21) Landmark Scenario 6:** John Husband has just completed his own self-response interview. His wife, Susan Wife, is not available to be interviewed, but John has agreed to serve as a proxy respondent for her. What do you do about landmark events at the start of the proxy interview? (Circle the best action.)

<p>a) Re-ask the "Landmarks" question and record new landmarks for Susan. Since John is serving as a proxy for his wife, ask him to "get inside Susan's head" and tell you about events during the past year that his wife would consider memorable. 11.5%</p>	<p>b) Don't re-ask the "Landmarks" question. Since it is John's own landmarks, not Susan's, that will help him remember events, ask him to help you recall his landmarks and when they happened, so you can enter them again in Susan's EHC interview. 20.8%</p>
<p>c) Re-ask the "Landmarks" question and record new landmarks for Susan. Ask John to tell you about memorable events during the last year that involved Susan. 60.2%</p>	<p>d) Don't re-ask the "Landmarks" question. Because it's a proxy interview, landmark events are irrelevant and unnecessary. 5.6%</p> <p>blank 1.1% multiple response 0.7%</p>

**PART 3. LANDMARK EVENT PROCEDURE QUESTIONS**

Select the best response for each question

22) What is the minimum number of landmark events that you must record for each respondent? (Circle only one)

<p>a) <i>One; you must record at least one landmark event for each respondent. Landmarks are essential to the EHC method; probe aggressively! If intensive probing fails, supply one yourself, using what you already know about the respondent and/or commonly understood 'generic' landmarks such as Easter, the 4<sup>th</sup> of July, Christmas, etc.</i> 10.8%</p>	<p>b) <i>Two; you must record at least two landmarks for every respondent. If the respondent has nothing to offer after you have read the 'Landmarks' question, supply two yourself, using what you already know about the respondent and/or commonly understood 'generic' landmarks such as Easter, the 4<sup>th</sup> of July, Christmas, etc.</i> 2.6%</p>
<p>c) <i>Three; you must record at least three landmark events for each respondent. Landmarks are essential to the EHC method; probe aggressively! Everyone has plenty of memorable events to report, if you dig hard enough.</i> 2.6%</p>	<p>d) <i>Zero; no one <u>must</u> produce any landmark event reports. If none are forthcoming after reading the question and following up with a gentle probe, move on.</i> 82.9%</p> <p>blank 0.7% multiple response 0.4%</p>

<p>23) What is the maximum number of landmark events you should record for each respondent? (Circle only one)</p>	
<p>a) <i>There is no set maximum; use your best judgement. Two or three good landmarks is plenty, especially if one of them is early in the reference year.</i> 71.8%</p>	<p>b) <i>There is no set maximum. Keep probing until the respondent tells you he cannot think of any more landmarks. Then probe at least one more time.</i> 23.1%</p>
<p>c) <i>Two. You should record no more than two landmarks. If the respondent reports more than two, use your judgement to select just the two that seem most memorable.</i> 0.4%</p>	<p>d) <i>Twelve. You should try to record one landmark for each month.</i> 3.0%</p> <p>blank 1.5% multiple response 0.4%</p>

<p>24) Why is it important to probe for landmark events that happened early in the year, if the respondent doesn't initially report any on his/her own? (Circle only one)</p>	
<p>a) <i>Early-in-the-year landmarks will be the most useful in helping respondents recall the hardest-to-recall events – those that happened the longest time ago before the interview.</i> 81.4%</p>	<p>b) <i>Your probing actions will make it clear to the respondent that landmark events are a crucial part of the EHC interviewing method.</i> 13.4%</p>
<p>c) <i>Landmark events will not be effective if they leave any holes or gaps in certain parts of the year or in certain seasons. Make sure they are spread out relatively evenly over the entire calendar year.</i> 3.4%</p>	<p>d) <i>It is NOT important to probe specifically for early-in-the-year landmark events. Landmarks from that period are unlikely to be very useful, because the respondent's memory for them is probably unreliable.</i> 0.7%</p> <p>blank 0.7% multiple response 0.4%</p>
<p>25) When should you give the respondent a calendar aid, and when you do so, what should you say? (Circle only one)</p>	

<p>a) <i>Using your best judgment about the timing, give the calendar aid (and a pencil, if necessary) to each respondent who is at least 21 years old, or who is responding as a proxy for another household member who is at least 21 years old. Say something like: "Many people find it helpful to look at something like this [hand calendar to the respondent] when they're trying to remember something."</i> 7.1%</p>	<p>b) <i>Give each new respondent a fresh calendar aid (and a pencil, if necessary) at the beginning of the EHC portion of the interview, along with a simple and brief explanation: "Here's something you might want to use to keep track of things as we go along." It is not necessary to give a fresh calendar aid to a respondent who, after his/her own self-response interview, is now serving as a proxy for another household member.</i> 63.9%</p>
<p>c) <i>Give each new respondent a fresh calendar aid (and a pencil, if necessary) at the very beginning of his/her interview, along with a brief explanation and request for consent: "I would like you to use this [hand calendar to the respondent] – it will come in handy later in the interview when I ask you about the months when things happened. Is that ok with you?" It is not necessary to give a fresh calendar aid to a respondent who, after his/her own self-response interview, is now serving as a proxy for another household member.</i> 16.0%</p>	<p>d) <i>Give the calendar aid (and a pencil, if necessary) to the respondent – self-responding or proxy – as soon as he/she shows any evidence of having difficulty recalling some event. (The calendar aid is a memory aid, so if there is no evidence of any difficulty then there is no need to bother with the aid.) Say something like: "Many people find it helpful to look at something like this [hand calendar to the respondent] when they're trying to remember something."</i> 11.5%</p> <p><i>blank 0.4% multiple response 1.1 %</i></p>

## PART 4. EVENT HISTORY CALENDAR SCENARIOS

*For questions 26 through 29 consider the current view of the EHC screen and the background information provided to determine what to do next in the following EHC scenarios*

- Press Enter to go into a period.

- Press Shift-Tab to return to the previous field without entering the period.

Topic	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Landmark events												
Residences												
Mental History												
Mom												
Dad												
Enrollment												
Job												
Job2												
Job3												
Job4												
Job5												
Job6												
NoJob												
SSI												
FS												
Food Stamp												
TANF												
GenAssis												
WIC												
Private1												
Private2												
Medicare												
Medicaid												
Military												
OtherCoverage												
NoCoverage												

**Questions**

**FS**

Food Stamp receipt?  1 Yes  2 No  3 Refused

Period:  0-New

Prior to January, what MONTH did you first start receiving Food Stamps continuously?

Prior to January, what YEAR did you first start receiving Food Stamps continuously?

Now I'm going to ask questions about the months of receipt that you just reported for the year of 2008. What changed in your life to cause you to need assistance?

New Child (or other dependent) or pregnancy.

Set Period: From:  To:

Not sure  1-Not sure

Current view of the EHC screen for Question 26

Background information for Question 26	
<p><b>Landmarks</b> MAR - visited sister in CA AUG - daughter moved to NC for school</p>	<p><b>Residences</b> no change, lived at current address for many years</p>
<p><b>Marital Status</b> widowed</p>	<p><b>Mom and Dad</b> did not live with either parent</p>
<p><b>Enrollment</b> none</p>	<p><b>Jobs</b> job 1: early MAY thru DEC job 2: JAN thru mid FEB no job: FEB thru APR</p>

**26) Event History Calendar Scenario 1:** In response to the Programs question, the respondent reports that she got Food Stamps in 2009. She remembers that they ended in May, as soon as she got her new job, but can't remember when the spell started. Based on the Event History Calendar information provided, what should you do next? *(Circle the best possible action.)*

<p>a) <i>Probe for the start of the Food Stamps spell using the date of the most memorable previously-reported event as a reference point: "So, your daughter went off to school in August. Do you remember how many months it was before that happened that the Food Stamps started? Six months? Seven months? Eight months?"</i> 4.5%</p>	<p>b) <i>Probe for the start of the Food Stamps spell using the dates of relevant previously-reported events as reference points. A good way to start would be: "Do you remember – did they start after you lost your job in mid-February?" If more probing is needed, a good follow-up would be: "Had the Food Stamps already started when you went to visit your sister in March, or did they start after you got back?"</i> 86.3%</p>
<p>c) <i>Don't probe any further at this point. The respondent doesn't know the start of the Food Stamps spell, so enter "don't know." There's a good chance that something later on in the interview will trigger her memory, in which case you will come back into this part of the EHC and enter the newly-recalled start date.</i> 1.5%</p>	<p>d) <i>People in financial need often receive benefits from more than one program at a time, so probe for receipt of other kinds of public assistance, and hope that those events trigger recall of the start of the Food Stamps spell: "When you lost your job did you receive other programs such as WIC, maybe, or TANF benefits, or something like that?"</i> 5.6%</p> <p><i>blank 1.1% multiple response 1.1%</i></p>

Current view of the EHC screen for Question 27

Topic	Category	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Landmarks	Landmark events												
Residences	Residence history												
Marital History	Marital history												
Mom	Mom												
Dad	Dad												
Enrollment	Education												
Job	Job/business												
Job2	Job/business												
Job3	Job/business												
Job4	Job/business												
Job5	Job/business												
Job6	Job/business												
NoJob	Job/business												
SSI	SSI program												
FS	Food Stamp												
TANF	TANF Program												
GenAssis	General Assistance												
WIC	WIC Program												
Private Insurance	Private insurance												
Medicare	Medicare insurance												
Medicaid	Medicaid insurance												
Military	Military insurance												
Other coverage	Other sources?												
NoCoverage	Anytime no coverage?												

- Use the text box to make a selection.
- Type '1' for Yes, '2' for No, or '3' for Refused.
- If a selection has been made, press Enter to continue.
- Choosing 'yes' will bring you to the 'period' question. Otherwise, you will be taken back to the 'List of Topics'.
- Press Shift-Tab to return to the 'List of Topics' without making a selection.

List of Topics

Landmarks
Residences
Marital History
Mom
Dad
Enrollment
Job
Job2
Job3
Job4
Job5
Job6
NoJob
SSI
FS
TANF
GenAssis
WIC
Private2
Medicare
Medicaid
Military
Other Coverage
NoCoverage

Questions

Private1

Private health insurance?

1) Yes  2) No  3) Refused

Self Period

From:  To:

1) Not sure  2) Not sure

Period  0-New

Background information for Question 27	
<u>Landmarks</u> JUN - son's HS graduation AUG - trip to beach	<u>Residences</u> moved into current residence in AUG; previous residence was JAN through JUL
<u>Marital Status</u> married	<u>Mom and Dad</u> No
<u>Enrollment</u> None	<u>Jobs</u> job1 - SEP thru DEC no job - JAN thru AUG
<u>Programs</u> Food Stamps - FEB thru MAY; TANF - FEB thru MAY General Assistance - JUN thru AUG	

27) **Event History Calendar Scenario 2:** The FR asks: ... At any time between January and December of 2009 were you covered by a private health insurance plan? The respondent replies: I don't think so, no. I don't think I had anything. Based on the Event History Calendar information provided, what should you do next? (Circle the best possible action.)

a) Probe about health insurance, either through Medicaid while respondent was receiving public assistance, or through the new job: "Just to be sure – your new job didn't provide any health coverage, is that correct? [Respondent: "That's right"] And what about before you got your job, when you were getting various kinds of public assistance. Were you covered by Medicaid during any of that time?" <p style="text-align: right;">81.4%</p>	b) Probe about the start and stop dates of the General Assistance benefits, since the respondent has reported dates for that program that aren't in line with the others: "Are you sure you got the General Assistance <u>after</u> your son graduated, and not before?" Or: "Are you sure about when you got the General Assistance? Did you get that while you were living here, or while you were in your previous house?" <p style="text-align: right;">5.2%</p>
c) Move on to the next topic, following the instrument path. There are no "red flags" here. There is no evidence that anything important might have been missed, and no evidence that the respondent is having any recall difficulties, so no special probing is called for. <p style="text-align: right;">11.5%</p>	d) One person in the household (the son) was clearly enrolled in school during the first part of the year, so maybe others were, too. Using what you know about her son's enrollment, probe to make sure the respondent's report of no enrollment is correct: "Your son was enrolled in school before he graduated. Are you sure you weren't enrolled at any time during the year?" <p style="text-align: right;">0.0%</p> <p>blank 1.1% multiple response 0.7%</p>



Background information for Question 28	
<u>Landmarks</u> FEB - won \$500 in lottery JUN - promotion at work OCT - garage burned down	<u>Residences</u> no change; lived at current address for many years
<u>Marital Status</u> married	<u>Mom and Dad</u> No
<u>Enrollment</u> JAN thru JUN and SEP thru DEC; part-time	<u>Jobs</u> job1 - JAN thru MAY job2 - JUN thru DEC; same employer, new job title, new duties/responsibilities, higher salary
<u>Programs</u> None	<u>Health Insurance</u> insured through employer all year; no other coverage

**28) Event History Calendar Scenario 3:** Based on the Event History Calendar information provided, what should you do next? *(Circle the best possible action.)*

a) <i>A job change is often associated with a change in health insurance coverage; a loss of coverage between jobs is quite common. Probe to make sure there was no period of being uninsured between job1 and job2: "What about when you got that promotion in June? Did you have the same coverage after that as you'd had before? Or was there any time, even for a day or two, when you weren't covered at all?"</i> 22.3%	b) <i>Urge the respondent to mark the calendar aid to indicate health insurance coverage for the entire year: "Do you see the 'Health Insurance' line of that calendar I gave you? Why don't you mark a line through the whole year to show you were covered all year." (If you haven't already done so, have him/her do the same thing on the 'Residences' line.)</i> 6.0%
c) <i>Using a landmark event, probe to make sure that the respondent is recalling the date of the promotion correctly: "Just to be sure: Did you get that promotion before or after your garage burned down?"</i> 3.0%	d) <i>Follow the instrument path. Things make sense, and the respondent is showing no apparent recall problems, so no special probing is called for.</i> 66.5%  <i>blank 1.5% multiple response 0.7%</i>

Current view of the EHC screen for Question 29

**Event History Calendar ver 2.21**  
 Browse Topics F3-Check Progress F10-Exit EHC

- Press Enter to go into a period.  
 - Press Shift-Tab to return to the previous field without entering the period.

Topic	Category	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Landmarks	Landmark events												
Residences	Residence history												
Mental History	Mental history												
Men	Men												
Dad	Dad												
Enrollment	Education												
Job	Job/business												
Job2	Job/business												
Job3	Job/business												
Job4	Job/business												
Job5	Job/business												
Job6	Job/business												
NoJob	NoJob												
SSI	SSI program												
FS	Food Stamp												
TANF	TANF Program												
GenAssst	General Assistance												
WIC	WIC Program												
Private1	Private Insurance												
Private2	Private Insurance												
Medicare	Medicare Insurance												
Medicaid	Medicaid Insurance												
Military	Military Insurance												
OtherCoverage	Other sources?												
NoCoverage	Aviame no coverage?												

**Questions**

**Job**

Work Job  
 Yes  No  Refused

Period  New

Set Period  
 From  To     
 Not sure  Not sure

Do Katie Doe work for an employer, or was Katie Doe self-employed, or did Katie Doe have some other work arrangement?  
 help job  
 was self-employed  
 other work arrangement  
 Don't Know  
 Refuse

00000023 EHC\_Intro 2:40:16 PM 10-16-2009 Context Respondent: Katie Doe P9015232

Background information for Question 29	
<p><b>Landmarks</b> (Same as Scenario #3)            FEB - won \$500 in lottery;            JUN - promotion at work            OCT - garage burned down</p>	<p><b>Residences</b>            no change;            lived at current address for many years</p>
<p><b>Marital Status</b>            Never Married</p>	<p><b>Mom and Dad</b>            Lived with Mom and Dad all year</p>
<p><b>Enrollment</b>            JAN thru JUN</p>	<p><b>Jobs</b>            job 1 - [Proxy respondent is unsure when the job started] thru DEC</p>

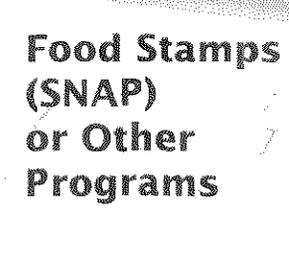
29) Event History Calendar Scenario 4: This is a proxy interview with the same respondent as in Event History Calendar Scenario 3 above. When asked about Katie's employment in 2009, the respondent replied, "I know her job ended in December, but I can't remember when it started." Based on the additional Event History Calendar information provided, what should you do next? (Circle the best possible action.)	
<p>a) <i>Don't worry about the missing start date; people forget things all the time. The respondent doesn't remember the date so enter 'Don't Know' and move on to the next topic.</i> 4.8%</p>	<p>b) <i>Use your judgment and make the best guess you can. In this case, the respondent probably started the job after the end of the school term, so put the beginning of the job1 spell in June.</i> 1.1%</p>
<p>c) <i>Probe for the start of the job1 spell using dates of previously-reported events as a reference point, e.g.: "Do you remember whether she was still in school when she started her new job, or was it after the term ended?"</i> 85.1%</p>	<p>d) <i>Probe neutrally about the start of the job1 spell, urging the respondent to think as hard as he/she can: "It's important to know when that job started. Please think very carefully. When did it start? Take your time..."</i> 3.4%</p> <p><i>blank</i> 5.2% <i>multiple response</i> 0.4%</p>

**ATTACHMENT B**

2010 SIPP-EHC Field Test – Calendar Aid

Event History Calendar

2009

Notes	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
 <p>Life Events</p>												
 <p>Housing</p>												
 <p>School</p>												
 <p>Work</p>												
 <p>Food Stamps (SNAP) or Other Programs</p>												
 <p>Health Insurance</p>												

**Calendario de la Historia  
de los Acontecimientos  
en su Vida**

**2009**

Notas	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
<b>Acontecimientos Importantes en su Vida</b>												
<b>Vivienda</b>												
<b>Escuela/ Universidad</b>												
<b>Empleo</b>												
<b>Cupones de Alimentos (SNAP) u Otros Beneficios</b>												
<b>Seguro de Salud</b>												