

**Web Development in the PSID:
Translation & Testing of a Web Version
of 2015 PSID Telephone Instrument**

June 21, 2017

Katherine A. McGonagle

Vicki A. Freedman

Jamie Griffin

Institute for Social Research, University of Michigan

Acknowledgements: Funding was provided for this research from the National Science Foundation (NSF 1157698). We gratefully acknowledge the contributions of PSID and SRO Staff: Mary Dascola (IDPM Team Leader), Rachel LeClere (SMP GSRA), Youhong Liu (SRO Programmer), Meredith House (SRO Survey Director), April Beaulé (PSID Data Processing Team Leader), and Mohammad Mushtaq (PSID Applications Development Team Leader).

Contents

STUDY AIMS	3
BACKGROUND AND OVERVIEW	3
METHODS.....	4
RESULTS	6
1. Eligibility and Response Rates	6
2. Response Rates by Stated Mode Preference	6
3. Timing of Accessing Web Instrument	7
4. Browser Types and Operating Systems.....	7
5. Instrument Length and Pauses	7
6. Session-level Paradata: Number of sessions, days to complete, back-ups, version buttons, consistency checks	8
7. Assistive Documents	9
8. Comparison of Key Estimates between 2015 PSID telephone interview and All-Stars Web	10
9. Modularizing Experiment.....	13
10. Qualitative Feedback from PSID All-Star Respondents	14
SUMMARY.....	15
Table 1. Sample Characteristics of All-Stars & Core PSID.....	17
Table 2. Eligibility and Response Rates Overall and by Mode Preference	17
Table 3. Browser Types and Operating Systems	18
Table 4. CATI & Web Timings for Comparable Sections; Web Actual & Self-Report for Entire Instrument.....	18
Table 5. Number of sessions, browsers, back-ups, version buttons, and consistency checks.....	18
Table 6. Document Types Reported by Document Users in All-Stars.....	19
Table 7. Comparison of Key Assets and Debts	20
Table 8. Means of K6 items, overall score, and high distress (N=94)	20
Table 9. Length and days to complete by instrument version.	22

STUDY AIMS

During the spring of 2014 the PSID study team began to explore options for integrating web interviewing into the Core PSID interview, conducted primarily by telephone since 1972. A key motivation was ensuring that PSID continued to be a premier source of data for the social and behavioral sciences as the study approached its 50th anniversary. Adding web as a data collection mode could potentially help the study maintain its high response rates and data quality.

Some evidence on the feasibility of these options for large-scale studies was beginning to emerge from the American Community Survey (US Census Bureau 2012) and the UK sister study, Understanding Society (e.g., Couper 2012; Jäckle et al. 2015). In 2014 PSID also began to explore internet options for data collection. With support from NIA (P01 AG029409 and R01 AG040213), PSID successfully conducted mixed mode (web, paper) supplements with PSID respondents, achieving response rates of 67% in 2014 and 81% in 2016, respectively (weighted). In addition, with pilot support from the Michigan Center on the Demography of Aging (P30 AG012846) PSID successfully tested in a laboratory context several issues related to translation of the core telephone instrument to an online format including use of instructions/probes and use of unfolding brackets following "don't know" responses. The pilot work revealed a number of additional design issues that would need to be addressed in the translation of the PSID instrument to web.

After a comprehensive review of the research literature on web design, the conclusion of the PSID investigators and the PSID Board of Overseers at this time was that more evidence was needed to understand the potential effects of offering web as part of a mixed mode design on PSID's response rates, data quality, and costs. With support from the National Science Foundation, PSID obtained a supplemental grant (NSF 1157698) in July 2014 to examine evaluate the use of web as a viable complement to telephone interviewing in the next cycle while minimizing risks to panel continuity. Specifically, the supplement had several aims:

- 1) Translate PSID's computer-assisted telephone instrument (CATI) to self-administered web, and create technical specifications and program the 2015 Core instrument for web;
- 2) Test the web program in 2015-2016 with two samples: a) a sample of approximately 1,000 families from an external (i.e., non-PSID) internet panel to confirm that the instrument was functioning as intended and to evaluate competing question design features (randomly assigned to respondents); and, b) a supplemental sample of highly cooperative respondents from the PSID representing those likely to complete their interview online without excessive follow-up.
- 3) Evaluate operational issues, data quality and – to the extent possible – cost. Operational issues of interest included obtaining an estimate of instrument length, number of sessions needed to complete, whether respondents had technical problems (e.g., problems logging in, etc). Data quality assessments included comparisons between 2015 Core PSID and the web instrument on distributions of key estimates.

BACKGROUND AND OVERVIEW

The goal of this project was to develop a web-based Core PSID instrument and assess the feasibility of offering it to PSID families during regular biennial data collection in the next funding cycle as a complement to telephone. The project was conducted over two main phases: Phase 1 began in September 2014 and included the translation of questions from the 2015 Core PSID computer-assisted

telephone interview (CATI) to a self-administered web instrument, the specification of these questions in Box and Arrow format for programming into Blaise, and the test of the instrument on an external convenience sample (“PSID Opt-In Test”). Beginning in January 2016, Phase 2 activities consisted of reprogramming and testing the instrument with a sample of cooperative PSID families (“PSID All-Stars Test”).

During Phase 1, from September 2014 through February 2015, PSID staff developed research-based visual design guidelines and a set of standardized question templates as the foundation of converting questions from interviewer-administered telephone to self-administered web. By reviewing each question in the instrument and applying 1 of 30 web-conversion templates to each question, a complete translation of the entire Core PSID instrument was made, with the exception of questions from the sections on New Births (J) and New Head/Spouse-Partner (K/L) since the initial instrument was designed for “reinterview” respondents with no between-wave family composition changes. The translation of questions to web was based on the principle of maintaining consistency as much as possible with the question wording in CATI (although there are some exceptions as described below, including the minor use of grids and differences in the employment sections between web and CATI).

In February of 2015 a complete web specification of the 2015 PSID questionnaire was provided to Blaise programmers in the Survey Research Operations (SRO) group at the Survey Research Center and programmed and extensively tested over a six-month period. Only questions that asked about household head (i.e., not questions about spouse/partner and other family unit members) were programmed in this first phase. Two versions of the instrument were developed in order to carry out experimental tests of specific design features. During the summer of 2015, each instrument version was randomly assigned to two groups of respondents purchased from an external web-panel (i.e., convenience sample) with overall demographic characteristics that closely matched those of PSID household heads and 960 completed interviews were obtained. Over the time period between August 2015 and December 2015, the team undertook data processing and analysis to assess which design features tested in the two instrument versions were preferable.

A second phase of development began in January 2016 that included the reprogramming and testing of the instrument to include the optimal design features, and the addition of question content for spouses/partners and other family unit members. Additional work focused on the development of a new series of questions designed to screen for family composition changes, the incorporation of survey data preloaded from 2015 Core PSID for dependent interviewing, the design and programming of the web portal, welcome and thank you screens, and the development of data collection protocols.

This report provides a detailed summary of the results of the second phase test of the administration of the PSID web instrument to a sample of cooperative PSID families – the PSID All-Stars Test.

METHODS

Sample Selection. The sample frame for the PSID All-Stars test was comprised of respondents to the 2016 PSID Wellbeing Mixed Mode Study (“WB”; NIA R01 AG040213), which actively collected data for approximately 16 weeks via web and paper from March 2016 through the end of June 2016 (and then continued to accept responses through the end of the year). The WB study was limited to heads and spouses/partners in PSID families who were aged 30 years and older in order to avoid over-burdening respondents who would be participating in the upcoming 2017 PSID Transition into Adulthood Supplement.

PSID All-Star respondents were identified according to the following criteria:

1. The individual completed WB within 4 weeks of launch (i.e.g, required minimal follow-up to complete);
2. The All Star respondent was the PSID respondent to the 2015 Core interview;
3. The family unit of the All-star respondent did not have a boyfriend/girlfriend or husband of head in 2013 Core or 2015 Core interviews;
4. The housing unit of the All-Star respondent did not include multiple family units in 2013 Core or 2015 Core; and
5. The 2013 Core family unit of the All-Star respondents included five or fewer family members.

Criteria 2-4 were met by 4,069 individuals and 2,285 of them responded to WB on the web within 4 weeks. The pool from which the All-Stars sample was drawn consisted of these 2,285 individuals. Two replicates of 100 randomly selected respondents were drawn from the All-Stars pool and invited to participate in the study, one in June 2016 and a second in September 2016.

As shown in Table 1, the sample characteristics for the 200 individuals invited to participate in All-Stars and the 2,285 respondents from which they were drawn differed from the overall PSID sample in multiple ways. All-Stars were overrepresented in the SRC sample, were significantly older on average (by definition since only individuals aged 30 years and older were included), had higher education and income and were more likely to be married. On average, All-Stars respondents required 14 calls to complete their telephone interview during 2015 compared to 22 calls on average for all PSID families, making them a high priority to move to web for saving field resources. No significant differences between the 200 randomly selected individuals and the larger pool of All-Stars were observed.

Sample Eligibility. Respondents who had any family composition changes since the 2013 PSID interview were screened out at the beginning of the All-Stars test. This decision was based on pilot testing (NIA P30 AG012846) finding that respondents had difficulty correctly completing the complex PSID family roster as a self-administered instrument. Completed at the start of the interview, the content of the family roster determines the instrument pathing (i.e., which questions about family members should be asked), and populates question text using the names, ages, and relationships of rostered family members. Thus, to minimize the potential for error in the instrument, a series of screening questions assessing family composition change since the 2013 Core PSID interview (i.e., did any family members move out, die, give birth/new father/adopt, change marital status, move back in after being away, and new family members including new births) was included at the beginning of the instrument. Respondents endorsing any change were flagged as ineligible, taken to an exit screen, and told that a check for \$20 would be sent to them in appreciation of their willingness to assist the project.

Contact Protocol. Invitations in the form of a letter from the Study Director requesting help testing an online version of the Family Economic Study (FES) interview and a Study Information sheet were sent via regular postal mail (See Appendix). The invitation to Replicate 1 was sent June 27, 2016, and the invitation to Replicate 2 was sent September 7, 2017. The invitation letter was slightly modified in Replicate 2 to alert respondents that they may not be eligible and if so would receive \$20 for their willingness to assist. The invitation provided the url address of the web survey and unique login credentials, described the voluntary nature of the request and confidentiality protections, and included an offer of \$100 upon completion of the interview. Two reminders containing the web address and login credentials were sent via regular postal mail 10 days after initial invitation likely arrived (July 15 for Replicate 1 and September 17 for Replicate 2) and day 20 (July 25 for Replicate 1 and September 27 for Replicate 2). In order to assess how easily these “All-Star” respondents (who would likely represent the

lowest-effort web respondents across all PSID families) would take up the invitation, no additional nonresponse reminders were sent.

Instrument. The technical systems used in the All-Stars web test were Blaise 5.0.v964 for the survey questions and the Michigan Survey Management System (MSMS) for the sample management system. The web interview included questions from 9 of 11 sections of the 2015 Core PSID instrument, excluding the sections on new births and new head/spouse-partner background. The instrument included all the questions asked in the Core about heads, spouses/partners and other family unit members, in contrast to the Phase 1 Opt-In test which asked questions only about heads. Preload information (except for the family roster which used data from the 2013 Core PSID) used the same structure and contained most of the same variables as the preload for the 2015 PSID Core computer-assisted telephone interview (CATI). Prior wave information about education, employment, and pensions was preloaded for verification into the web instrument.

Interview Instructions and Access. Upon logging into the interview, respondents were reminded of the confidentiality protections, told that the interview works best when completed on a computer, notebook or tablet (rather than a smartphone), that each session would time out after 15 minutes of inactivity, and that they could return to the interview by entering the login number contained in the mailing.

Access to the interview was blocked for individuals using smartphones and small mobile devices with screen sizes below 9.5 inches (and iPad users which is an issue being worked out since these screens are 9.7 inches).

The All-Stars study was reviewed by the University of Michigan Institutional Review Board and received the determination of Not Regulated (HUM00101072).

RESULTS

All results are collapsed across the replicates since the replicates were randomly drawn from the All-Stars pool and there are no significant differences between the replicates in any of the outcomes reported below. One case of 99 completed interviews had unreliable paradata and is excluded from all timing and paradata analyses. Various analyses exclude other outliers as described.

1. Eligibility and Response Rates

Overall, 80.0% of the 200 individuals (n=160) who were invited to the study attempted to log in to the portal (Table 2). Thirty-eight percent (n=61) of these individuals were not eligible to complete the interview due to one or more of the family composition changes described above. All of the remaining 99 eligible cases provided an interview: 96 provided complete interviews and 3 provided acceptable partial interviews (AAPOR RR#2 = 71.2%). There were zero instances of starting the interview and breaking off permanently.

2. Response Rates by Stated Mode Preference

Mode preference for completing the FES interview was obtained through a question asked in the 2015 Core PSID interview ("Suppose we could offer you the option of completing this interview over the internet rather than by telephone. If you could choose how to complete your interview, would your first choice be internet, telephone, or would you not have a preference?"). Rates of attempting to complete the interview and completing the interview were highest among those preferring internet (85.9% and 78.6%), followed closely by those preferring telephone (78.3% and 76.2%). Nearly two-thirds of those with no stated

preference attempted to complete the interview (65.3%) and half successfully completed the interview (50.0%).

3. Timing of Accessing Web Instrument

All-Stars respondents generally attempted to access the web instrument soon after being invited. About one-third of the 200 individuals invited to the study attempted the survey before the first reminder was mailed on day 10, and another 26% attempted before the second reminder was mailed on day 20. Thus, more than half (57%) attempted the web instrument with either no reminder or only 1 reminder.

4. Browser Types and Operating Systems

More than half of all respondents used Chrome (55.4%), followed by Internet Explorer (23.2%), Firefox (12.5%) and Safari (8.9%; Table 3). The most frequently used operating system was Microsoft Windows, accounting for 85.0% of sessions, followed by Mac OS (11.4%) and Linux (3.6%). Only one respondent who took more than one session to complete their interview logged back in using a different browser than in the first session.

5. Instrument Length and Pauses

Overall instrument length was calculated using the nine sections of the questionnaire that were comparable between web and CATI. One exception is the employment section (BC/DE), which differed rather substantially between modes. CATI uses an unstructured interviewer-led conversational approach through the administration of an event history calendar (EHC) to obtain employment and residential information. During the translation to web, a deliberate decision was made to convert the information collected within the calendar to a series of independent questions (based on the approach used in CATI pre-EHC, 2001), due to the substantial development and testing needed to create a comparable web-based calendar. A second, minor exception is the limited use in web of grids (also known as matrix questions) where the rows are a set of items and the columns are the response options with the respondent selecting one response in each row. This is in contrast to the design in CATI where the interviewer would read the initial question ("lead-in") and the response options; for subsequent items the interviewer would optionally repeat the lead-in and response options. Grids were used for the 8-item series on t-2 income in section R and for the 6-item 30-day distress series in the health section.

Across the nine comparable sections of the interview, the average length of the All-Stars web instrument was about 82 minutes (~72 median minutes; Table 4) compared to about 72 minutes in CATI (~72 median minutes). On average, respondents took nearly 10 minutes (13.6%) longer to complete the 2015 instrument in web than CATI.

Five of nine sections took significantly longer to complete in web compared to CATI (housing, employment, food, income, health), 1 section was faster in web (philanthropy), and there was no difference between the modes for 3 of the sections (pensions, t-2 income, and wealth). The employment sections for head (BC) and spouse/partner (D/E; among the n=64 with spouse/partners) were the sections with the largest length differences between the modes, taking substantially more time to complete in web. Excluding the employment sections, Web remained significantly longer than CATI by 6.7 minutes (~63 minutes compared to ~56 minutes respectively, $p < .02$).

One possibility for the higher interview length of Web compared to CATI is that respondents are pausing or taking mini "breaks" as they complete the web instrument, in contrast to the experience in CATI where the interviewer keeps the respondent moving through the instrument generally without such breaks. Some support for this was found by examining time stamp data and identifying periods of time when respondents "paused" in the web instrument, with a pause defined as a period of time of at least 2

minutes and less than 15 minutes (at which point the instrument timed out). About 35% of respondents completed the web instrument without pausing at all. Among those who paused at least once (n=63, removing 1 outlier), the average number of times pausing was about 4, the median was 3, and the 75th percentile was 5. The average total pause duration (across all pauses) was about 16 minutes, the median was 11 minutes, and the 75th percentile 21 minutes. Thus, Web respondents seem to be taking very brief pauses of 3-5 minutes on average, which could account for some of the difference in overall instrument length between the modes.

Respondents were also asked to report how long they thought it took them to complete their web interview. As shown in the right-hand panel in Table 4, among 80 respondents answering this question, Web was reported to take about 82 minutes on average (median of 75 minutes), or approximately 11 minutes less than the actual time it took (93 minutes on average; as generated from time stamp data). Note that this comparison was based on the entire web instrument and not just the 9 sections included in the comparison with CATI, since self-reports are probably based on a mental calculation of the total experience of completing the interview, from logging on through the exit screen.

Respondents may perceive the duration of Web as less than actual duration because of the pauses they are taking as they complete the Web, which they do not count in their mental calculation of Web duration. Examination of the time stamp data show that there is a negative association between the number of pauses and the difference in perceived and actual web survey duration ($r = -0.55$, among respondents taking at least 1 pause (n=63)). That is, among respondents who take at least one pause, the more pauses taken, the smaller the difference between self-report and actual duration (i.e., greater accuracy in reported duration). We also found that those who reported using documents (e.g., tax returns, pay stubs) had significantly longer interviews (see Section 7), more pauses (about 4 pauses at 25 minutes total duration on average) and a longer total pause duration than those not reporting use (2 pauses at 15 minutes total duration on average), although tests of significance were not significant.

In summary, comparable sections of the 2015 instrument took significantly longer in Web than CATI; the employment section (which was structured differently between modes more than any other section) was especially longer in Web. Even when the employment section was excluded from the timings comparison, the instrument took about 12% longer in Web than CATI. About two-thirds of Web respondents took an average of 3-5 mini-breaks or pauses, lasting about 3-5 minutes each. Individuals using documents took significantly longer (see Section 7) on Web than CATI and had more pauses.

6. Session-level Paradata: Number of sessions, days to complete, back-ups, version buttons, consistency checks

6.1 Number of Sessions and Days to Complete: Respondents took an average of 1.7 sessions to complete the web instrument (Table 5). The majority completed the instrument in one session (59.2%) and nearly all respondents (91.8%) finished in 3 or fewer sessions. Nearly 80% of respondents completed the instrument in one day, and 90% completed the instrument across 2 days. The efficient completion of the web instrument by the All-Stars is encouraging and implies that few interviewer resources will be needed to encourage these respondents to finalize their survey.

6.2 Backing-Up: Respondents backed-up from one question to a previous question in the instrument an average of 11.6 times; more than half (52.1%) backed up at least 6 times. No single field accounted for the preponderance of backups; however, nearly one-third of backups (30.4%) occurred in Section G (income), 15.2% occurred in Section F (expenditures), and 11.2% occurred in Section BC (employment of head). Respondents reporting document use backed up significantly more times than did respondents reporting not using documents (18 times vs. 6 times, respectively, $p < .01$).

6.3 Unit of Time Version Buttons: Version buttons that permitted respondents to change the reported unit of time were provided for 43 items (in the sections on housing, employment, pensions, and health) that asked about expenditures over a unit of time (Figure 1; e.g., “About how much rent do you pay a month?”). These buttons were designed to replicate the approach taken in CATI which permits respondents to answer questions about expenditures using various time units. On average, respondents changed the unit of time on 2.3 version buttons; more than half of respondents (59.2%) used 2 or more of these buttons. Version buttons for questions on amount of time away from work and amount of pension contributions made by self and employer (BC10a – c, and P15, P18) comprised half of all the questions on which the unit of time was changed.

6.4 Consistency Checks: Visual design was used in the Web instrument to stand-in for the telephone interviewer who probes respondents when they provide incomplete or out of range answers. Three main types of consistency checks were added to the instrument.

First, “amount / per” checks were deployed if an amount was entered but the unit was missing (e.g. “How much do [you / you and your family living there] usually pay for electricity per month?”). This consistency check was added for a total of 69 questions, including 10 questions in the housing and utilities section (A), 7 on employment (BC/DE), 13 questions about expenditures (F), 21 on income (G), 1 about past year income (R), 2 on assets (W), 10 on pensions (P), and 5 on health (H).

Second, “amount range” checks were added when respondents entered out of range amounts (e.g. “How much did that alimony amount to in [PYEAR]?”). These checks were added to 17 questions, including 4 questions on income (G), 2 on pensions (P), and 11 on health (H).

Third, two types of “date range” checks were added to ensure that reports of when events started and ended did not occur after the current date (e.g., “In what month and year did the foreclosure start?” and “In what month and year did [YOU / [HEAD] / [WIFE/“WIFE”]] stop working for this employer?”). Range checks for start dates were added to 4 questions, including 2 questions on mortgages (A) and 2 on income (G), and checks for end dates were added to 4 questions, including 1 question on employment (BC/DE), and 3 on income (G).

Note that due to question pathing, some respondents may not have received every consistency check question. Future analysis may attempt to create case-level counts of version buttons and consistency checks to more precisely understand the circumstances under which they are deployed.

Consistency checks were triggered by about one-third of respondents (36.7%) who deployed 1.4 consistency checks on average. Of those deploying a consistency check, two-thirds deployed exactly 1, 29% deployed exactly 2, and 2 cases deployed 3 or 4 check. In sum, consistency checks were added to 94 questions in the instrument. The addition of consistency checks, although deployed for only one-third of respondents, appears to be an important feature for ensuring data quality.

7. Assistive Documents

The use of documents to assist in completion of the web interview was reported by more than half of respondents (54%; Table 6) and 71% of those reporting using at least one document used multiple documents. Of those using documents, 64.4% used a tax return, 48.9% used utility bills, 22.2% used spending records, 17.8% used their mortgage and/or property tax statements, and 15.6% used insurance statements (Table 6). Other documents used included pension statements, other bills, bank statements, medical bills, business addresses. Respondents using documents had actual web interview lengths approximately 24 minutes longer than those not using documents (112 minutes on average compared to 88 minutes on average, $p < .05$). Document use is not assessed in CATI. However, those using

documents in Web have average 2015 PSID CATI lengths significantly greater than those not using documents in Web (91 minutes on average compared to 82 minutes on average, $p < .05$), suggesting they likely also use documents in CATI. The influence of document use on interview length is greater in Web than in CATI, with an increase on average interview length of about 27% in Web compared to about 11% in CATI.

8. Comparison of Key Estimates between 2015 PSID telephone interview and All-Stars Web

Respondent reports for items obtained in the 2015 PSID interview were compared with those reported during All-Stars Web. Differences between these reports could reflect true changes that occurred between 2015 and 2016, the change in the reference period covered in the question, or mode effects (e.g., understanding the question differently when an interviewer is not present to clarify). We attempted to generate a “benchmark” of expected change by first comparing the consistency of reports in 2013 PSID CATI with 2015 PSID CATI for the All-Stars respondents. We then examined report consistency between 2015 PSID CATI and All-Stars Web. The main goal of the analysis was to document the general magnitude of consistency / differences being reported between CATI and Web.

8.1 Reports of Key Assets & Debts. For reporting of key assets and debts, the rates of consistency in “status” reported between 2013 and 2015 CATI waves were generally comparable to those between 2015 CATI and 2016 Web (Table 7). Reports of median values of various assets were slightly less consistent between waves of CATI than between 2015 CATI and 2016 Web, indicating that the passage of time may have had a greater influence on reports than the change in mode. Overall the magnitude of these differences in reports of status and value are low.

For example, 93.9% reported no change in home ownership/renting between 2013 and 2015 CATI waves, and 88.9% reported no change between 2015 CATI and Web. Among the 60 respondents who were homeowners and provided valid home values across all three waves (2013 and 2015 CATI and Web), the median change in home value was \$24,500 between 2013 and 2015 CATI waves, and \$0 between 2015 CATI and Web.

No change in whether there was a balance in various financial accounts was reported by 89.9% between 2013 and 2015 CATI waves, and by 84.9% between 2015 CATI and Web. Among the 67 respondents with money in such accounts across all three waves, the median change in account value was \$300 between 2013 and 2015 CATI waves and \$400 between 2015 CATI and Web. There was no change in whether there was credit or store card debt reported by 78.8% between 2013 and 2015 CATI waves and 79.8% between 2015 CATI and Web. Among the 25 respondents with such debt in all three waves, the median change in debt amounts was \$950 between 2013 and 2015 CATI waves and \$800 between 2015 CATI and Web.

Examination of data on private annuities and IRAs shows that 78.8% reported consistently on whether they had such accounts between 2013 and 2015 CATI waves and 85.9% reported consistently between 2015 CATI and Web. Among the 37 respondents with such accounts in all waves, 62.2% exhibited no change in whether the investment was invested mostly in stocks, mostly in interest-earning assets, or split between stocks and interest-earning assets between 2013 and 2015 waves and 64.9% reported no change between 2015 CATI and Web. Among the 31 respondents with such accounts in all three waves, the median change in account value was \$18,000 between 2013 and 2015 CATI waves and \$10,000 between 2015 CATI and Web.

8.2 Reports of 30-day Emotional Distress. The six-item K6 measure of non-specific emotional distress (Kessler et al. 2002) has been administered in Core PSID using CATI since 2001 and was included in the 2016 Well-being and Daily Life (WB) instrument from which the All-Stars sample was drawn, permitting an

examination of reports of distress across mode. Comparisons were made across the 2013 and 2015 waves (CATI), and the 2016 WB and All-Stars web on the means of each of the six items comprising the K6 measure, as well as on the K6 overall score (i.e., the average of the six items), and on the rate of “high distress” (i.e., K6 scores greater than or equal to 13).

Prior research leads us to expect higher reports of sensitive items in self-administered modes than interviewer-administered modes (e.g., Cernat et al. 2016; Tourangeau & Yan 2007). We examined the influence on reports of distress (self-administered versus interviewer-administered) of mode and context (how prior questions affect responses to subsequent questions). Mode effects were tested by comparing reports from the 2015 CATI (telephone) with both 2016 All-Stars Web and 2015 WB (web). Mode and context effects were tested by comparing 2015 CATI (the questions preceding 30-day emotional distress items were on physical health) with 2016 WB (the emotional distress questions were a continuation of the same content on psychological states); a comparison of 2016 WB with 2016 All-Stars report, which were both obtained via web, but differed in the preceding questions, allowing a ‘gross’ assessment of the effects of context.

Table 8 shows that reports of the individual items and overall score between the two waves of CATI are generally stable. Comparing All-Stars Web to 2015 CATI, we see that five of the six individual K6 items, the overall score and the percent with high K6 (only 1-3 individuals) are all higher in All-Stars, but the differences in only two items achieved statistical significance. With only 94 observations, statistical tests of significance are under-powered in this analysis. However, comparing 2016 WB with 2015 CATI and All-Stars provides suggestive evidence that individuals will report higher rates of sensitive items in a self-administered setting. Reports of distress from WB are significantly higher than from CATI (for all items, the overall score, and the percent high-K6). At first glance this may indicate a pure mode effect. However, a comparison of WB and All-Stars (same mode, relatively close in time) finds that three of six items and the overall score are significantly higher in WB. Such a finding is consistent with a question context or priming effect (e.g., Schwarz & Clore 1983) in which exposure to a stimulus activates a concept in memory that is given weight in subsequent judgement tasks. Notably, the vast majority of the WB content – including the content that immediately precedes the K6 questions - is focused on subjective wellbeing, whereas in All-Stars Web the preceding content focuses on physical health.

In summary, reporting differences between modes were reassuringly modest and depended on the content. The differences in reports of the ownership and amounts of key assets and debts between CATI and web were low. Differences in psychological distress also were modest, consistent with a context effect rather than mode effect.

8.3 Design & Testing of Open-Ended Questions on Occupational Detail for Web. Collection of information in surveys that include open-ended questions is facilitated by interviewer probing and follow-up questions. A key example is the collection of occupational information, which must be sufficiently detailed in order to apply widely used coding systems to the data, such as the 3-digit U.S. Census Occupation Classification Codes. In PSID, this information is collected in an open-ended format by interviewers who ask respondents to provide information about their major activities and duties, and then follow with additional questions and probes to ensure the responses are sufficiently detailed.

A key question is whether respondents are able to provide adequately detailed information on open-ended questions without the presence of an interviewer. A related question is whether there are mode effects – do reports provided through self-administration differ substantively from those provided via interviewer-administration?

A sub-project was undertaken to address these issues and optimize and test the design of open-ended questions on occupation detail during the web-translation of the telephone instrument. In the CATI version of the instrument, the interviewer asks a series of three separate questions (“What is your occupation?” “What sort of work do you do?” “What are your most important activities or duties?”), and then “probes” the respondent by repeatedly asking “anything else” until the respondent is unable to provide additional information. In the Web version of the instrument, the three occupation questions appeared individually and on separate screens. We investigated two issues: Did respondents provide sufficient detail for an expert coder to apply the 3-digit Census codes? Did the coder apply the same 3-digit code to the responses provided in CATI and Web?

Methods: Analyses comparing respondent occupational reports from the 2015 CATI instrument with those from the 2016 All Stars web instrument were based on individuals who did not experience a job change. There were 59 such respondents; of these, 39 also reported a current main job for their spouse/partner, yielding a sample of 98 current main jobs from Web that could be compared to 2015 CATI. Job title and occupation and industry descriptions from the three questions asked in Web were concatenated together into one text field in the coding application to make these data appear comparable to the CATI data. Web and CATI responses were added in separate rows and in a random order in the application (i.e., Web and CATI reports for an individual respondent did not appear sequentially) and provided to an expert coder. The coder was simply asked to provide a 3-digit Occupation code to the data, and was blind to which respondent provided the information and whether the mode was Web or CATI.

Results: We first compared the *quantity* of occupation information provided by calculating the total character count of the open-ended text fields. Nearly three times more characters were recorded by interviewers in 2015 CATI than were provided by respondents in 2016 All Stars web (CATI: mean=337.9 characters, SD=183.58, median=300.5; All Stars: mean=116.1 characters, SD=75.74, median=89; paired test statistic $t(97)=12.56, p<.0001$).

To compare the *quality* of occupation information provided, we examined whether the coder had sufficient information to assign a code, and the rate of agreement between the Web and CATI code for the same jobs. Results showed that although the web responses were significantly shorter than the CATI responses, the coder had sufficient information to assign a code to 100% of the occupational reports.

A simple rate of agreement defined as the percentage of cases with identical codes was calculated. Levels of agreement were generally high across modes: 87.8% of cases were assigned identical first digits; 79.6% were assigned identical first and second digits; and 73.5% were assigned identical first, second, and third digits.

Consistent with the findings for all 98 cases, among the 12 cases assigned different first digits, significantly more characters were recorded by interviewers in CATI than were provided by respondents in Web, and this difference was large (CATI: mean=374.7, SD=196.6, median=333.5; All Stars: mean=81.1, SD=29.2, median=81; paired test statistic $t(11)=5.11, p<.001$). On average, non-matches provided longer CATI reports (the difference is not statistically significant: $t(96)=0.74, p=0.5$) and shorter web reports ($t(39.57)=-3.33, p<.01$) than matches which had more equal reports between modes (CATI: mean=374.7 vs. mean=332.8, respectively; web: mean=81.1 vs. mean=120.9, respectively).

We also examined whether agreement between Web and CATI reports on the first digit of the two codes varied by self-report or proxy-report (i.e., spouse/partner occupation being reported by respondent). There was a higher rate of non-matches for proxy reports compared to self-reports (18.0% vs. 8.5%, respectively) but this difference was not statistically significant ($\chi^2(1)=1.961, p=0.2$).

Summary: PSID respondents provided significantly less detail on web than telephone, but the information was sufficient to achieve a 3-digit occupation code for all cases. Moreover, the occupation descriptions provided for web and telephone were consistent, leading to generally high matches on the final 3-digit code. Overall, the results indicate that open-ended information can be successfully collected through self-administered modes (and probing may be unnecessarily high in telephone).

9. Modularizing Experiment

We designed an experiment to test two methods of providing encouraging messages and reminders about incentives to motivate respondents to complete the lengthy PSID instrument. In long and complex interviewer-administered surveys such as the PSID, the interviewer plays this key role. An open question during the design of the All-Stars web instrument was whether respondents would be able to complete the 60-80 minutes self-administered interview without the encouragement of the interviewer. We considered the use of embedded progress bars that allow respondents to track their completion status but a review of the literature found mixed results about their benefits (e.g., Conrad et al. 2010).

Methods: Two versions of the All-Stars instrument were programmed that differed in the messages given to respondents about completion progress and incentives: a single-session version and a modularized version. Respondents were randomly assigned to one of the versions once they completed the household roster and their eligibility was confirmed on the basis of having no family composition changes. The initial messages were provided on the final landing page (after welcome, introduction, and the screening questions) prior to launching the instrument.

As shown in Figure 2, the landing page in the single session told respondents that the interview would take about 60-80 minutes (60 if the household had one family member and 80 if there were more than one) and that a check for \$100 would be sent once the survey was complete. After approximately one-third of the interview was completed – “Part 1” - (i.e., at the end of the employment section), respondents in this group were congratulated for finishing about one-third of the instrument and told they would receive a \$10 Amazon gift card in addition to the \$100 when they completed the survey. After approximately two-thirds of the interview was completed – “Part 2” - (i.e., at the end of the X section), congratulations were made for finishing about two-thirds of the instrument and respondents were told that both the gift card and the \$100 check would be provided upon completion of the interview. The messages after Part 1 and Part 2 were contained within the survey instrument, after the last question in each of the question sections, before the first questions in the subsequent sections began.

Respondents in the modularized version were told that the interview was divided into three parts (Figure 2), with each part lasting between 20-30 minutes (20 if the household had one family member and 30 if there were more than one), and that a check for \$100 would be provided upon completion of all three parts. The questionnaire content contained in each “part” was the same as in the single session. After Part 1 was completed, respondents exited from the questionnaire and were taken to a landing page where they were congratulated for finishing Part 1 and told that they would receive a \$10 Amazon gift card in addition to the \$100 when they completed the survey. After Part 2 was completed, respondents were again exited from the questionnaire and taken to a landing page where they were congratulated for finishing Part 2 and that both the gift card and the \$100 check would be provided upon completion of the interview.

Thus, there were two key differences between the single session version and the modularized version: 1) conveying the overall expected interview length in the single session versus how long each third of the interview would take in the modularized version; and 2) messages about progress and incentives were

conveyed within the instrument (single session) versus on a landing page after being exited from the questionnaire (modularized version). The amount of the incentives did not differ between the versions.

Research Questions: A key question was whether the encouraging messages about progress and incentives influenced rates of breaking off and interview length. In particular, those in the modularized version who exited from the questionnaire after each part may have used the opportunity to breakoff, whether temporarily and thus take more sessions or days to complete the interview, or permanently, resulting in lower response rates. Alternatively, exiting from the physical questionnaire after 20-30 minutes sessions may have provided a welcome “mini” break, and the tailored landing page may have increased the salience of the encouraging messages, resulting in a greater number of completed interviews for those in the modularized version. Those in the single session group received no such respite and had messages embedded within the questionnaire that may have been less salient, potentially leading to breakoffs due to fatigue. It is also possible that respondents preferred to finish the instrument in one time block, instead of being forced to exit and reenter the questionnaire in two blocks.

Results: There were no differences between the two groups in how long it took them to complete the All-Stars web survey or the 2015 PSID CATI survey or how long they perceived the web survey took. There were also no differences in the number of sessions over which respondents completed the All-Stars web survey.

On average, respondents in the modularized version started the survey significantly earlier in the field period than did respondents in the single session (14th day in field vs. 19th day in field; $p=.0363$). Respondents in the modularized version also completed the web survey over significantly more days than respondents in the single session version (2.5 days vs. 1.1 days, respectively; $p=.0149$).

Summary: The modularized version had the positive benefit of reducing the number of days it took for respondents to start the survey compared to the single session, yet once they started, respondents in this group took slightly more sessions and significantly more days on average to complete the survey. This is consistent with the idea that breaking down the overall interview minutes may make the task seem more palatable and encourage respondents to get started, but being forced to exit the questionnaire may also encourage breakoffs, leaving the survey unfinished over more days. Delays starting the survey, as well as delays finishing the survey, could each increase the use of resources to remind respondents to complete the survey. While the small number of cases assigned to each group yielded underpowered statistical tests, the results are suggestive and call for additional testing.

10. Qualitative Feedback from PSID All-Star Respondents

A brief series of questions at the end of the interview asked respondents to rate their experience completing the web instrument across a variety of domains. Respondents were strongly positive about their experience completing the web instrument. Of those responding to the debriefing questions, the vast majority strongly agreed or agreed with the following statements: “I felt comfortable completing the instrument (98.8%),” “error messages were easy to understand (100%),” “it was easy to work out how to fix problems (96.2%),” “the survey was easy to complete (94.0%),” and “the look of the survey was pleasant (93.9%).”

Respondents were asked to provide their preference for completing the FES interview in the future: 97% chose web and 3% chose telephone. More than half of those preferring web (57.5%) cited reasons such as web “let me complete the interview at my own pace...within my schedule... and over multiple sessions if needed.” Completing the interview “faster” was mentioned by 17.8% (reflecting the perception that web was faster, despite it taking objectively longer than CATI). About 12% mentioned that completing the

interview on web allowed them to check their answers using documents; another 12% liked being able to complete the interview themselves and not have to talk with an interviewer on the telephone.

SUMMARY

The overall goals of this project to develop a web-based Core PSID instrument and assess the feasibility of offering it to PSID families during regular biennial data collection as an alternative to telephone were achieved. A self-administered web version of the 2015 PSID telephone instrument was successfully translated and tested with an external convenience sample, reprogrammed to optimize question design, and re-tested on a sample of 200 cooperative respondents from PSID families. A series of data analyses were conducted to provide information about potential savings of interviewer resources; respondent burden; and data quality. These results help define the next steps for broader implementation of web during regular data collection.

This test indicates that interviewer resources would be saved among web-friendly families with characteristics matching those drawn from the pool of All-Stars. About 80% of these families logged in to attempt the interview and 100% of those who were eligible completed the interview. Minimal nonresponse follow-up efforts would be needed for such families who predominantly completed the survey in a single session with no reminder or only 1 reminder. Thus, the pool from which the All-Stars were drawn – approximately 80% of the 23% of all PSID families selected into the All-Stars pool (~1840 PSID families) – can be expected to embrace the offer to complete their interview on web, and do so with minimal field resources. Saving 14 calls across 1840 families would result in a reduction of about 26,000 calls – or 13% of all calls (~200,000) made to complete interviews across the field period, allowing more resources to be devoted to difficult cases. To better understand the cost efficiencies of web, a subsequent test should broaden the pool of respondents such as by including additional cases that completed the Wellbeing web module over a longer time period.

The fact that about 38% of PSID respondents have a between-wave family composition change that would make them ineligible to complete the web instrument as currently programmed is a major caveat to these promising resource savings, and highlights the need to develop a user-friendly self-administered version of the household roster as a next step in web development.

Respondent burden in terms of actual time to complete the web instrument may not be lower than CATI. In fact, All-Stars respondents took about 10 minutes (~14%) longer to complete the instrument in web than CATI, with the employment sections taking substantially more time to complete. The reason for the longer employment section is not understood and needs additional exploration. Perceived burden may be lower, however, as respondents experienced the web survey as faster than CATI which could increase their enjoyment and/or willingness to complete a web interview. This is consistent with the strongly positive reviews of the web experience reported nearly unanimously by respondents.

Comparisons between estimates obtained from web with those from the 2015 CATI provide generally reassuring information about web data quality and mode effects. The comparisons provide an imprecise test, however, since the reasons for changes are unknown, and may be due to real change, the passage of time, or mode effects. Overall, these “stable” families had comparable consistency in reports of assets and debts (both “status” and amounts) between CATI waves 2013 and 2015 with those reported between 2015 CATI and web. Rates of emotional distress reported in web were higher than in CATI, consistent with other research finding that self-administered modes elicit greater endorsement of sensitive items compared to interviewer-administered modes. In this regard, web may be a preferred mode for collecting sensitive information.

In summary, there are several next steps in the progression toward offering web during regular data collection. As already described, a user-friendly household roster needs to be developed to accommodate the substantial number of PSID households who have a family composition change. The types of families that we engage in testing needs to be expanded to include less “web-friendly” families (i.e., those who took longer to complete PSID mixed mode supplements) and younger adults who have so far not been included in tests. Finally, nonresponse contact protocols need to be developed. The small test comparing the single session version with a modularized version holds some promise for future development in showing that offering the instrument in short blocks may reduce the number of days for respondents to start the survey, but has the tradeoff of taking more sessions and days on average to complete the survey. Expanding the test bed of PSID families in web development may also necessitate the development of targeted contact protocols.

Table 1. Sample Characteristics of All-Stars & Core PSID

		PSID 2013 (N=9063)	Invited All- Stars
Sample Type			
	SRC	60.1	76.5
	SEO	32.4	20.0
	Immigrant	7.5	3.5
Age			
	average	44.7	47.5
	median	42.0	45.0
	% 30s	23.8	31.5
	% 40s	16.7	19.5
	% 50s	18.4	21.0
	% 60s	11.8	13.5
Gender			
	% female	61.1	57.5
	% male	38.9	42.5
Completed Education (years)			
	average	13.6	14.9
	median	13.0	16.0
Couple Status			
	% single	46.3	31.5
	% coupled	53.7	68.5
Total Family Income			
	average	\$ 68,168	\$ 124,054
	median	\$ 48,377	\$ 79,736

Table 2. Eligibility and Response Rates Overall and by Mode Preference

Group	Number Invited	Number Attempting	Attempting Response Rate	Number Eligible	Number Complete	Interview Response Rate
Total	200	160	80.0%	139	99	71.2%
<u>By Mode Preference</u>						
Internet	128	110	85.9%	84	66	78.6%
Telephone	23	18	78.3%	21	16	76.2%
None	49	32	65.3%	34	17	50.0%

Table 3. Browser Types and Operating Systems

Browser Types				
	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Chrome	93	55.4	93	55.4
Firefox	21	12.5	114	67.9
Internet Explorer	39	23.2	153	91.1
Safari	15	8.9	168	100.0
Operating Systems				
Linux	6	3.6	6	3.6
Mac OS	19	11.4	25	15.0
Windows	142	85.0	167	100.0
Frequency Missing = 1				

Table 4. CATI & Web Timings for Comparable Sections; Web Actual & Self-Report for Entire Instrument

Comparable Sections				Entire Web Instrument		
minutes	CATI*	Web (Actual)¹	Difference	Web (Actual)²	Web (Self-Report)	Difference (self-report minus actual)
mean	71.9	81.7	9.8**	93.0	81.9	-13.6
median	72.5	73.8	1.3	83.4	75.0	-10.6
<i>N</i>	95	95	95	95	80	80

¹Includes only questionnaire sections A, B, C, D, E, F, G, H, M, P, R, W

²Includes entire instrument from login through exit

**p<.01

Table 5. Number of sessions, browsers, back-ups, version buttons, and consistency checks

	Number of Sessions	Number of Browsers	Number of Backups	Number of Version Buttons Selected	Number of Consistency Checks Deployed
mean	1.71	1.01	11.62	2.28	0.52
median	1.00	1.00	6.00	2.00	0
<i>N</i>	98	98	98	98	98

Figure 1. Example of a Unit of Time Version Button

A31, A31YR. About how much rent do you pay a month [including rent for the lot]?		
	To enter an amount per year instead, click here.	
\$	<input type="text" value="1 - 99,999"/>	.00 per month
	To enter an amount per month instead, click here.	
\$	<input type="text" value="1 - 99,999"/>	.00 per year

Table 6. Document Types Reported by Document Users in All-Stars

Document types reported in Web	%(N=45)
Tax returns	64.4
Utility bills	48.9
Spending records/family ledger/budgeting software/checkbook	22.2
Mortgage statement/Property tax statements	17.8
Insurance bills	15.6
Pay stubs	8.9
Retirement-pension statements	8.9
Other bills	8.9
Bank statements	6.7
Medical bills	4.4
Business addresses	2.2

Table 7. Comparison of Key Assets and Debts

	% Same Status ^a			Median Change in Value (\$) ^b			% Same Investment Vehicle ^c		
	2013 CATI to 2015 CATI	2015 CATI to 2016 Web	N	2013 CATI to 2015 CATI	2015 CATI to 2016 Web	N	2013 CATI to 2015 CATI	2015 CATI to 2016 Web	N
Home ownership	93.9	88.9	99	24,500	0	60		n/a	
Cash accounts ^d	89.9	84.9	99	300	400	67		n/a	
Debts	78.8	79.8	99	950	800	25		n/a	
Private annuities or IRAs	78.8	85.9	99	18,000	10,000	31	62.2	64.9	37

^aHome ownership: no change in home ownership/rentership; Cash accounts: No change in whether had any money in various accounts; Debts: No change in whether had credit or store card debt; Private annuities or IRAs: No change in whether had money in private annuities or IRAs; ^bAmong those with such assets/debts in all three years and providing valid values in all three years; ^cWhether invested mostly in stocks, mostly in interest-earning assets, or split between stocks and interest-earning assets; ^dChecking or savings accounts, money market funds, certificates of deposit, government bonds, treasury bills

Table 8. Means of K6 items, overall score, and high distress (N=94)

	(1) 2013 CATI	(2) 2015 CATI	(3) 2016 Wellbeing Web	(4) 2016 AllStars Web	Significance of (3) – (2) mode + context effect	Significance of (4) – (2) mode effect	Significance of (3) – (4) context effect
30-day distress							
Individual items ¹							
Sad	0.27	0.24	0.50	0.46	***	**	ns
Nervous	0.89	0.68	1.11	0.84	***	ns	**
Restless	0.94	0.76	1.16	0.72	**	ns	***
Hopeless	0.14	0.14	0.46	0.33	***	*	ns
Everything an effort	0.52	0.45	0.98	0.61	***	ns	***
Worthless	0.13	0.13	0.34	0.21	**	ns	ns
Overall score ²	2.88	2.39	4.54	3.17	***	ns	***
%High K6 (>=13)	2.13	1.06	5.32	3.19	*	ns	ns

¹0=None of the time, 1=A little of the time, 2=Some of the time, 3=Most of the time, 4=All of the time

²Minimum score=0, maximum score=24

Significance of mean differences estimated using Wilcoxon signed rank test

Significance of difference in % high K6 estimated using McNemar’s exact test

p*<.05, *p*<.01, ****p*<.001, ns=not significant

Figure 2. Design of Version Experiment

	Single-Session		Modularized	
	Location:	Message:	Location:	Message:
Interview Length	Landing Page	This interview takes about [<i>1 FU MEMBER: 60 / >1 FU MEMBER: 80</i>] minutes to complete.] You will receive a check for \$100 once you complete the interview. Thank you again for your help!	Landing Page	We have divided the interview into 3 parts. Each part takes about [<i>1 FU MEMBER: 20 / >1 FU MEMBER: 30</i>] minutes to complete. You will receive a check for \$100 once you complete all 3 parts. Thank you again for your help!
Consent and Starting	Landing Page	The questions asked in this online version will be very familiar to you since they are identical to the questions asked during our telephone interview in 2015.	Landing Page	The questions asked in this online version will be very familiar to you since they are identical to the questions asked during our telephone interview in 2015.
		This interview is completely voluntary and confidential. The answers that you give will be kept confidential to the maximum extent allowable under federal and state law. Click here to start		This interview is completely voluntary and confidential. The answers that you give will be kept confidential to the maximum extent allowable under federal and state law. Part 1 Click here to start Part 2 Part 3
After 1/3 Interview completed:	Last question in BC/DE	You have finished about one-third of the interview. Congratulations!	Exit to Landing Page	You have finished Part 1. Congratulations!
		To show our thanks, you will receive a \$10 Amazon gift card in addition to the \$100 check when you complete the interview. Click here to continue.		To show our thanks, you will receive a \$10 Amazon gift card in addition to the \$100 check when you complete all three parts. Part 1 Part 2 Click here to start Part 3
After 2/3 Interview Completed:	Last question in X	You have finished about two-thirds of the interview. Congratulations!	Exit to Landing Page	You have finished Part 2. Congratulations!
		You will receive the gift card and \$100 check when you complete the interview. Click here to continue		You will receive the gift card and \$100 check when you complete all three parts. Part 1 Part 2 Part 3 Click here to start

Table 9. Length and days to complete by instrument version.

	Instrument Version	
	Single Session	Modularized
Outcome	Mean	Mean
Web length	98.0	95.7
Web length-self report	85.5	84.4
2015 PSID CATI length	87.1	87.1
Number of days before starting	18.8	14.1*
Number of sessions	1.5	1.9
Number of days to complete	1.1	3.5*

*p<=.05

There were n=57 respondents assigned to the modularized group and n=41 assigned to the single session. For the analysis involving self-reported web length, data are available from n=35 assigned to the single session and n=47 from the modularized group.