UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

January 17, 2013

2013 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS13-RER-02

DSSD 2012 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS12-MP-07

MEMORANDUM FOR	ACS Research and Evaluation Advisory Group
From:	Anthony G. Tersine ***/Signed/*** Assistant Division Chief, Decennial Statistical Studies Division
Prepared by:	Rachel Horwitz ACS Data Collection Methods Branch Decennial Statistical Studies Division
Subject:	Using Paradata to Identify Potential Issues and Trends in the American Community Survey Internet Instrument

Attached is the final American Community Survey Research and Evaluation report "Using Paradata to Identify Potential Issues and Trends in the American Community Survey Internet Instrument." This report summarizes the results of the paradata analysis using data from the November 2011 American Community Survey Internet Test. The objective of the analysis was to determine whether the issues and respondent behaviors identified in the paradata analysis from the April 2011 American Community Survey Internet Test still exist.

If you have any questions about this report, please contact Rachel Horwitz at 301-763-2834 or Jennifer Tancreto at 301-763-4250.

Attachment

cc:

ACS Research and Evaluation Workgroup

Debbie Griffin	(ACSO)
Todd Hughes	
Agnes Kee	
Andrew Roberts	
Brian Wilson	
Dave Raglin	
Sandra Clark	
Samantha Fish	
Stephanie Baumgardner	
Kathy Ashenfelter	(CSM)
Temika Holland	
Beth Nichols	
Victor Quach	
Barbara O'Hare	(DIR)
Nancy Bates	
Frank Vitrano	
Enrique Lamas	
Mary Ann Chapin	(DMD)
Justin McLaughlin	
Tony Tersine	(DSSD)
Michael Bentley	
Mary Davis	
Steven Hefter	
Joan Hill	
Samantha Stokes	
Jennifer Tancreto	
Mary Frances Zelenak	
David Johnson	(SEHSD)
Bob Kominski	
Colleen Hughes	(POP)
Anne Ross	
John Studds	(ASD)
Glenn Eanes	
Joe Mistichelli	

American Community Survey Research and Evaluation Program
January 17, 2013

Using Paradata to Identify Potential Issues and Trends in the American Community Survey Internet Instrument

FINAL REPORT

Rachel Horwitz, Jennifer Guarino Tancreto,
Mary Frances Zelenak, Mary Davis
Decennial Statistical Studies Division

United States™ Census Bureau

Intentionally Blank

TABLE OF CONTENTS

LIST OF TABLES AND FIGURE

EXECUTIVE SUMMARY

Objective

This report provides results of the paradata analysis from the November 2011 American Community Survey (ACS) Internet Test. This analysis focuses on screens and behaviors that were identified as problematic in the April 2011 ACS Internet Test to determine whether these issues still exist. Additionally, we examine whether the respondent interaction with the instrument is similar to what we expect using the paradata analysis from the 2011 April Internet Test as a reference.

Methodology

Paradata were collected as a by-product of the Internet data collection process for the November 2011 ACS Internet Test for all Internet respondents. That test included five experimental treatments and two strata with a total sample size of 100,000. Four of the experimental treatments used in the Internet Test were not considered for the paradata analysis; only paradata from the Push Internet on an Accelerated Mailing Schedule with New Reminder Postcard treatment were analyzed to provide insight into the effectiveness of the Internet survey instrument for all respondents that chose that mode of response. We limited the analysis to this treatment because it is the treatment selected for 2013 ACS production operations. A total of 5,291 Internet responses were the focus of this analysis and strata results were collapsed. Weights were incorporated into all estimates (other than completion times and devices used) to account for probability of selection into the sample.

Research Questions and Results

Although this analysis focuses on behaviors and issues uncovered in the paradata analysis from the April 2011 ACS Internet Test, we cannot directly compare the April and November Test results because the treatment universes were different and weights were not used in the April analysis. However, all of the analyses for the November test yielded results that were within an expected range from what we saw in April and we did not uncover any new issues.

a. Are there any problematic screens or questions?

Overall, only 14 percent of people who accessed the survey broke off before completing it. The breakoffs occurred on over 100 different pages throughout the instrument, suggesting most of the breakoffs are not related to a specific screen or question, but rather the respondent needing or choosing to leave the instrument for another reason. We found a relatively high percent of breakoffs on transitional screens.

About 50 percent of respondents received at least one error message while taking the survey. The Place of Birth question rendered the highest number of error messages because respondents selected a radio button but did not provide the required write-in

of country or selection of a state from the drop-down box. A similar format appears multiple times throughout the survey, so the error rate on this particular screen may be inflated because it is the first time respondents see it. Additional research into different ways to present all of the questions with a radio button/write-in format could make it easier for respondents to answer and reduce the total number of error messages and the percent of respondents who receive an error message.

Respondents requested help on about 97 percent of the screens where help was available. A higher proportion of respondents requested help on the ancestry screen (13.1 percent), than on the other screens, and relative to the number of visits to the screen, the ancestry screen also had the greatest rate of help requests (5.8 percent)[1]. This suggests the concept of ancestry, independent of the mode of administration, may be difficult for respondents to comprehend. Additional research could uncover how to make the question easier to understand.

b. Do respondents use the Help option and does it appear to be useful?

Respondents did use the help link and it seems to have helped them answer questions. Overall, help was requested at least once by 38.5 percent of respondents. In 54.8 percent of cases, the information in the help link appears to have been used to generate an answer (no response option had previously been selected). Alternatively, only 5.1 percent of help requests resulted in a changed answer. This could either indicate that the respondent already had the correct response and the help verified that choice, or the help did not provide the information they needed to justify changing their answer.

c. Are there any problems with the re-entry authentication procedures?

The majority of respondents did not have any problems with the authentication procedures to re-enter the survey after either logging out, timing out, or exiting. Only 7.0 percent of the logins that followed a logout or timeout were invalid. However, about 60 percent of these cases never gained access to the survey. Therefore, experimenting with different types of authentication procedures may provide a way for these respondents to re-enter the instrument.

d. Is there any additional information about the use of the instrument or survey features the paradata can help explain?

In general, respondents interacted with the instrument as we expected. The majority used a computer, but 3.6 percent used a tablet and 0.9 percent used a phone to access the survey. Additionally, 80 percent of respondents completed the survey in one session. Finally, across all household sizes, respondents took 37 minutes to complete the survey, on average.

We should continue monitoring the use of mobile devices to determine whether we need to develop a mobile-friendly instrument moving forward.

[1] The difference between the percent of help clicks on the Ancestry screen and the next highest screen was significant using a one-sided test at the $\alpha = 0.05$ level.

1. BACKGROUND

Paradata are all of the data collected during the response process that do not include the response itself; or process data (Couper, 1998). They were used in the evaluation of the April 2011 ACS Internet Test instrument to identify potentially problematic questions or screens and to understand how respondents interact with the instrument. We identified several issues with the instrument including how often the error messages are displayed, their content and relatively frequent breakoffs on transitional screens and screens asking for sensitive information (Horwitz *et al.*, 2012). Additionally, using paradata, we were able to assess what types of devices participants used to access the survey, how many sessions it took to complete the survey on average, how frequently they requested help and on which pages, and what links and features they used.

For additional information on prior research using paradata in Internet surveys, please see "Use of Paradata to Assess the Quality and Functionality of the American Community Survey Internet Instrument" (Horwitz *et al.*, 2012).

Examining the paradata from the November 2011 ACS Internet Test will allow us to determine if the issues and behaviors identified in the April paradata analysis are likely to be trends that we will continue to see as the Internet mode moves into production.

2. FUTURE RESEARCH CRITERIA

We expect that this analysis will yield similar results as those seen in the April 2011 analysis. Based on this expectation, we defined the following criteria that would determine the need for future research:

- If issues we believed were problematic in the April analysis still exist, we recommend that future research be undertaken to solve these issues.
- If issues arise that were not identified in the April analysis, we will continue to monitor the issues as the Internet mode moves into production and will recommend research for solutions if the trend continues.

3. METHODOLOGY

3.1 Description of the ACS

The ACS is a mixed-mode, mandatory household survey of all persons living or staying at a sampled address. The Census Bureau samples about 3.54 million housing unit addresses for the ACS each year. From 2005 through 2012, most sampled units received a questionnaire in the mail. If they did not complete and return the questionnaire, they were switched to a Computer Assisted Telephone Interview (CATI) if we had a phone number for the address. A sample of the addresses that had still not completed the survey were then visited by a field representative to conduct a Computer Assisted Personal Interview (CAPI).

The ACS consists of a rostering section in which the respondent provides names for everyone living in the household, a series of demographic questions that are asked one question at a time for each household member, a series of questions about the housing unit, and finally detailed questions that are asked for each person in the household, one person at a time. At the time of the test, the average time to complete the survey in a production environment was 38 minutes. The actual length is dependent on the household size and the number of questions that are applicable, based on skip patterns.

3.2 Data Collection Methods

In the November 2011 ACS Internet Test, we tested five different strategies for notifying sampled units about the Internet response mode, but only one is used in this analysis. This report focuses on the Push Internet on an Accelerated Mailing Schedule with New Reminder Postcard treatment because it is the method that the Census Bureau selected to use in ACS production in 2013. Sampled units were notified using combinations of five ACS mailing pieces (pre-notice letter, initial questionnaire mailing, reminder postcard, and for nonrespondents only, replacement questionnaire mailing, and additional reminder postcard). These units received the paper questionnaire several weeks after the invitation to complete the survey online. For more information on the notification strategies and results, see "2011 American Community Survey Internet Tests: Results from the Second Test in November 2011" (Matthews *et al.*, 2012).

3.3 Sample Design

In the November 2011 ACS Internet Test, we stratified tracts into the same two groups used in the April 2011 test: Targeted and Not Targeted. The Targeted group consists of tracts containing households that we expect to use the Internet at a higher rate based on past research. The balance of tracts was placed into the Not Targeted group. At the time of sample selection, we suspected these groups varied by age, education, and computer experience and we wanted the opportunity to identify differences in responding behavior across these groups. For more information on the Targeted and Not Targeted groups, see Tancreto *et al*. (2012b).

We weighted the estimates and proportions in this report using the Internet Test base weights, which account for selection into the ACS and into each stratum[2].

3.4 Analysis Design

Included in this report is an analysis of the paradata collected during the ACS Internet Test in November 2011, which is intended to help assess the quality and efficiency of the Internet instrument. There were three major categories of paradata collected: survey access/authentication, session information, and features. The list of paradata that were used in this analysis can be found in the Appendix. The goal of this analysis was to study

[2] The devices estimates are not weighted because the devices are not linked to a specific household.

behavior related to the user's interaction with the Internet instrument to determine whether this behavior is consistent with the findings from the April paradata analysis. Paradata were collected and analyzed for every screen and for every household in the Push Internet on an Accelerated Mailing Schedule with New Reminder Postcard treatment that logged into the survey. This analysis focuses on observing response behaviors that stood out in the April paradata analysis, which used all treatments, such as where breakoffs and errors occurred, as well as general behaviors that explain how respondents use the instrument.

3.5 Internet Instrument Design

The ACS Internet instrument did not change between the April and November tests. The instrument was designed to be similar to the mail and CATI/CAPI ACS data collection modes. This strategy was used to minimize mode effects, while taking advantage of the technology, as used in CATI and CAPI, to improve data quality. Consistent with the paper and CATI/CAPI versions, the Internet instrument had four sections of questions (as described in Section 3.1). Unique to the Internet mode, at the end of the survey, the respondent had the option of reviewing responses or submitting the survey without reviewing. If respondents chose to review, they could select whether they wanted to see the housing data or the person data for each individual. In other words, they did not need to review all the data. Additionally, they could link to specific questions within the review to change their answers. For more information on the instrument design, please see Tancreto *et al.* (2012a).

4. LIMITATIONS

All of the limitations discussed in this section have a mild impact on the estimates presented in this report. However, they are not so problematic as to jeopardize the relevance of the findings. These are just caveats to keep in mind while reading the results because the numbers presented are not as exact as they may appear. Unfortunately, most of these issues concern the nature of paradata and cannot be fixed for future tests.

4.1 Server-side versus Client-side

Timestamp data are collected both on the client side (user) and the server side (Census Bureau). When respondents log in, log out, and submit their surveys, the time is captured on the server side. Timestamps for all other actions, such as when the respondent enters a page and selects an answer, are captured on the client side. For respondents on the East Coast who have their clocks set closely to the server's clock, this is not an issue. However, for all other respondents, their data points are not sequential. For example, because the login is on the server side, it will appear that respondents on the West Coast logged in to the survey three hours after answering the first question. Since these data points do not map to the order in which respondents progressed through the instrument, it can be difficult to identify whether a respondent logged out or left the survey by closing their browser.

Therefore, we could not use these server-side time stamps to calculate completion times. Rather than timing when a respondent logged in and logged out, we instead measured the time when they entered the first screen (which asks the respondent to verify his/her address) to when they entered the Presummary[3] screen because both of these measures are on the client side. However, these time stamps were subject to load time issues which will be discussed in Section 4.2.2.

4.2 Glitches

The use of JavaScript leads to several types of glitches in the collection of paradata. These glitches affect most of the estimates presented in this report. These glitches are not related to the instrument, but are simply due to the way paradata are captured and cannot be fixed. Additionally, it is not clear how big of a problem these different issues are. The glitches seem to appear randomly throughout the survey, so it is not possible to search for them explicitly. Most were identified while looking at other data anomalies. However, we do not believe they are very prominent.

4.2.1 JavaScript disabled

If a respondent does not have JavaScript enabled, no data will be collected from the user side. These individuals are included in the breakoff rate calculation, but are otherwise excluded from this analysis[4]. At least five households (out of 5,921 households that accessed the survey) did not have JavaScript enabled for the entire survey. It is very difficult to search for these households, so it is possible this limitation applies to more than the five identified households. It is possible households who have JavaScript disabled are different than other households.

4.2.2 Load times

Another potential issue with JavaScript is load time. If load times are long on a particular page, paradata may not be collected for that page. For example, in order to first log into the survey, respondents must first enter a User ID and then click "Login." Every respondent should then be presented with the "Address" screen asking if he/she lives at a specific address. However, there were several cases identified for which the "Address" screen does not appear in a respondent's paradata. Some respondents appear to enter the survey at the PIN screen, which appears three screens after the login, while some enter at a random page later in the survey. This could be either the result of a respondent enabling JavaScript part way through the survey, or long load times. Due to this anomaly, these cases had to be excluded from all duration calculations because it was not possible to determine exactly how long they had been in different sections of the instrument. Further, we are likely underestimating the exact number of respondents who accessed different screens, especially at the beginning of the instrument.

[3] The Presummary screen appears when the respondent has viewed all questions for every person on the roster. On this screen, the respondent is given the choice of reviewing his/her answers or submitting the survey.

[4] All logins and completions are logged on the server side and are not dependent on the user's settings.

Similarly, long load times can result in "holes" in the paradata. Specifically, several respondents skipped from the "Race" question to "Year Built," which is two questions later in the instrument. Further, we know these are issues with the paradata and not with the instrument because their final data output contains answers for the items missing in the paradata, which means the respondent saw the screen. These types of glitches cannot be fixed on the client side, but are problematic when trying to determine how many people visited each screen or what actions were taken on a screen because they are not included.

4.2.3 Timestamps

The program the Census Bureau uses to output Javascript, XML, collects paradata by nesting actions within headings. These paradata can then be sorted by any one of the variables corresponding to the headings. We opted to have the paradata sorted by time so we could see how respondents move through the instrument. However, the time is recorded in seconds. Therefore, it is possible for two actions to occur within the same second. This mostly occurs with write-in fields. Specifically, the timestamp for the write-in is not recorded until the respondent clicks out of the text box. Often, their next click is on the "Next" button. Therefore, they receive the same timestamp for both actions. When this occurs, Javascript does not know in what order to put these two events since it can only sort on one variable. Therefore, it selects one to be first and one to be second, which does not necessarily reflect the correct order. This makes programming a challenge because the paradata may show a respondent entering one page before ever leaving the prior. We attempted to account for various paradata orders, but some may have been missed, which would lead to an underestimate of statistics such as percentage of changes after requesting help.

4.3 Comparison across Tests

There are two factors that limit the comparison of data between the April 2011 Internet test and the November 2011 Internet test. First, some Internet respondents who had not completed their online survey did not receive a replacement questionnaire package in the April test that should have. In the November test, everyone who should have received a replacement questionnaire package did. This correction likely affected response rates and breakoff rates because in November, Internet respondents were reminded to complete their survey. Secondly, this report only contains data from one push panel, whereas the April paradata report included data across all four panels, two of which provided respondents a concurrent option to respond by paper or Internet. This could affect the characteristics of people who responded via the Internet.

Additionally, we did not conduct any statistical analyses across tests. These comparisons cannot be made because the paradata results from the April test were not weighted. Therefore, we are looking for similarities across tests and are not making direct comparisons. Where we note differences, they may or may not be statistically significant, but we think they deserve more attention.

5. RESEARCH QUESTIONS AND RESULTS

Due to the limitations discussed in Section 4.3, the analyses discussed in this section only make broad, indirect comparisons to the results from the April paradata analysis. Across all measures, we did not see any unexpected results that are cause for concern.

5.1 Are there any problematic screens or questions?

We identified several potentially problematic screens from the paradata analysis in the April test. This analysis will help us determine if respondents left the survey, received error messages, or requested help on similar screens as in the April test. If this is the case, it could suggest problems with the questions that we can research further to keep people in the instrument and moving without interruption.

5.1.1 Breakoff Rates

A breakoff is often defined as a case where someone begins a survey but does not finish it. For the purposes of this paradata analysis, we defined a breakoff as any respondent that accessed the survey (saw the first screen that appears after logging in), but did not reach the presummary screen, which appears after the respondent has seen all applicable screens for all people in the household. In total, 5,291 people accessed the instrument and 743 respondents broke off before completing the survey, resulting in an overall breakoff rate of 14.0 percent.

As expected, we also found that the breakoff rate increased as household size increased (Table 1). We saw a similar trend in the April analysis and in prior research because increases in household size lead to more questions and longer response time, providing more opportunity for the respondent to grow frustrated or be distracted from the survey (Horwitz *et al.*, 2012 and Galesic and Bosnjak, 2009).

Table 1. Breakoff Rate by Household Size

Household size	Number of households	Breakoff Rate (se)
1 person	40	3.6 (0.6)
2-3 people	338	11.8 (0.7)
4-5 people	166	16.2 (1.3)
6+ people	39	25.7 (4.5)

Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

The breakoff rate suggests most respondents complete the survey. However, for those that do not, it is important to know how far they progressed before leaving and on which screen they left. This information tells us how much data we are able to collect for these participants (breakoffs early in the instrument are more concerning than late breakoffs) and it can also tell us whether certain screens are triggers for breakoffs because they are transitional or sensitive in nature (Peytchev 2009).

We chose to investigate breakoffs using two different rates. The screen breakoff rate compares the number of breakoffs on each screen to the total number of visits to that screen, which tells us the screens with the highest rates of breakoffs. To determine where in the instrument the majority of breakoffs occurred, we also calculated the percent of breakoffs by comparing the total number of breakoffs on each screen to the sum of all breakoffs in the instrument.

Breakoffs are spread across most of the screens in the instrument. Respondents broke off at least once on about 71 percent of the screens in the instrument. Of the screens on which respondents broke off, about 88 percent had fewer than 10 of the 743 breakoffs. This distribution suggests respondents usually break off when they are either bored with the instrument, get distracted, or need to focus their time on something else, but likely not because they have a problem with a specific question.

There are, however, a handful of screens where respondents tended to leave the instrument more often than others. Table 2 provides screen breakoff rates for screens where the percent of breakoffs is greater than 2.0 percent for screens with a large number of screen visits. A review of the data across all screens found that no screen had a breakoff rate that exceeded 1.1 percent.

Table 2. Breakoff Rates by Screen

Breakoff Page	Breakoff Frequency	Number of Screen Visits	Screen Breakoff Rate (se)	Percent of Breakoffs (se)
Pick Next Person	130	11,899	1.1 (0.1)	17.5 (1.5)
Respondent Name	36	5,236	0.7 (0.1)	4.7 (0.8)
Amount Received in Wages	36	6,701	0.6 (0.1)	4.9 (0.8)
Finished Person	24	7,159	0.4 (0.0)	3.3 (0.7)
Date of Birth	25	13,202	0.2 (0.1)	3.1 (0.7)

Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

It is clear from Table 2 that the Pick Next Person screen accounts for the largest percent of breakoffs, although compared to the number of visits to the screen, the percentage is quite low. Additionally, the Finished Person screen accounts for 3.3 percent of breakoffs. As discussed in the April paradata analysis, these screens are both transitional (the screen provides information and there is no survey task to complete) and provide an easy stopping point for respondents. While the Pick Next Person screen was designed to provide flexibility to respondents with more than one household member, it may be worth experimenting with combining this screen with the Finished Person screen to see the impact on breakoffs.

In the April Internet Test, we also identified the PIN screen as a transitional screen that accounted for a relatively high percent of breakoffs (3.9 percent of breakoffs and 0.6 percent of screen visits). However, we do not see a similar result in the November test where the screen had a screen breakoff rate of 0.3 and accounted for 2.2 percent of all

breakoffs. It is not clear why breakoffs on the PIN screen appear lower, but it could have something to do with Push respondents as compared to Choice respondents in the April test. We should continue to analyze breakoffs on the PIN screen moving forward to determine whether this lower rate is a trend or an anomaly.

While we suspect the breakoffs on the Pick Next Person and Finished Person screens are a result of the transitional nature of those screens, we suspect the breakoffs on Respondent Name and Date of Birth are a result of the sensitivity of the question[5]. These screens ask personal and identifying information that some people may simply not want to provide. Across all screens other than Date of Birth, we saw similar results in April, suggesting this is a pattern we will continue to see moving into a production environment. The good news is that while breakoffs are occurring on these screens, they account for a very small percent of visits to the screen. On the other hand, they appear at the beginning of the instrument, so we collect very little information on these households. Providing information on why we are asking these questions may make some respondents feel more comfortable answering the question.

5.1.2 Error messages

All of the screens containing questions deemed critical for the ACS, questions critical for skip patterns, and all questions asking for a dollar amount have error messages associated with them. For example, if the respondent left one of the questions deemed critical completely blank, they received a message saying: "Please answer this important question." On the other hand, if a respondent enters an invalid value, they receive a message saying: "Please enter only numbers." In addition, there are more specific errors that instruct respondents on how to fix invalid entries which vary by question.

In total, there were 7,264 error messages rendered throughout the instrument by 2,603 respondents across 55 of the 56 questions with error messages[6]. On some screens, multiple error messages can appear at once, depending on what information the respondent provides. These error messages appear in multiple text boxes, however they just as easily could have been provided in one. The current design inflates the number of error messages that are rendered. Therefore, this analysis eliminates the duplicate messages and focuses on the number of times at least one error message was rendered. After taking the duplication into account, there were 4,828 total error messages rendered.

Overall, there were about 18 screens on which an error message was rendered in more than one percent of visits. Across all screens on which respondents received an error message, this percent of screen visits ranges from 27.5 percent to 0.1 percent. Table 3 shows five screens with among the highest percent of error messages rendered. The "Percent of Error Messages" column in Table 3 shows the total number of error messages rendered on each page as a percent of the total number of error messages rendered throughout the instrument (4,828). The "Percent of Screen Visits" column represents the

[5] Both of the sensitive questions had error messages associated with them if the question was left blank. Therefore, some of the respondents may have thought they could not continue without providing the information. Had they known it was a soft edit, they may have continued without answering the question.
[6] An error message was possible on the amount paid for second mortgage question, but it was not rendered.

number of respondents who received an error message on each page as a percent of the total number of times each screen was visited. Table 3 shows that about 1.8 percent of all the screens with a possible error message that were visited resulted in the display of an error message.

Table 3. Screens on which Respondents Frequently Received an Error Message

Page Error Rendered	Number of Error Messages Rendered	Number of Screen Visits	Percent of Error Messages (se)	Percent of Screen Visits (se)
Total	4,828	263,745	-	1.8 (0.1)
Place of Birth	1,186	12,011	24.8 (0.7)	9.9 (0.4)
Amount Received in Wages	434	6,701	9.0 (0.5)	6.4 (0.4)
Respondent Name	256	5,236	5.3 (0.4)	4.9 (0.3)
Year Built	223	5,134	4.7 (0.3)	4.3 (0.3)
Date of Birth	462	13,202	9.2 (0.5)	3.4 (0.2)

Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

The first thing to note from Table 3 is that these are the same screens that had a screen error rate of more than four percent in the April paradata analysis. Therefore, there is nothing alarming in these findings and no additional cause for concern.

It is also clear from Table 3 that the Place of Birth screen is responsible for many of the errors. As discussed in the April analysis, many of the errors on this screen and the Year Built screen involve clicking a radio button but failing to fill in the write-in information (Horwitz *et al.*, 2012). While we do not see a negative effect of respondents receiving this message because almost all respondents corrected the mistake (Horwitz *et al.*, 2012), many respondents clearly do not understand what is expected of them for these questions. Therefore, we should test different formats of the questions to make them more user-friendly.

The other screens that appear in Table 3 ask sensitive information so it is possible respondents tried to leave the question blank or did not know the information for all household members. These questions are much more difficult to change as compared to Place of Birth and Year Built. Specifically, for Place of Birth and Year Built, we only need to direct the respondent to the write-in box using an arrow, highlighting or bolding. To address more sensitive questions we would likely need to change the wording of the question (such as an unfolding bracket design for Amount Received in Wages where respondents are first asked if their wages were above or below a specified amount and then are asked more specific amounts or to write in an amount) and to explain why the information is necessary and how we plan to use it. However, we should continue to monitor these screens during production to ensure the error rate does not increase[7].

[7] The format of the Date of Birth question has been changed from a write-in to drop downs for production. This change should eliminate errors due to invalid entries for year of birth, especially.

5.1.3 Use of the Help feature

A Help link was available for 107 questions (out of 145 total questions) throughout the instrument[8]. Table 4 shows screens where Help was selected in at least two percent of visits and accounted for at least one percent of help requests[9]. The "Percent of all Help Clicks" column compares the total number of times help was requested on each screen to the total number of times help was selected in the instrument. The "Question Help Rate" compares the number of times help was clicked on each screen to the total number of visits to the screen (e.g., 5.8 percent of respondents who saw the Ancestry screen requested help). In total, only one percent of all screen visits resulted in a help request (for screens on which help was available). Additionally, the Question Help Rate had a fairly even distribution across screens: 55.1 percent of the screens with help available have a Question Help Rate of less than one percent.

Table 4. Screens on which Help was Most Frequently Selected

Page	Percent of all Help Clicks (se)	Question Help Rate (se)
Total	-	1.0 (0.0)[10]
Amount Paid in Real Estate Taxes	4.9 (0.4)	6.6 (0.5)
Ancestry	13.1 (0.6)	5.8 (0.3)
Pay for Electricity	5.0 (0.3)	5.1 (0.4)
Property Insurance	3.4 (0.3)	4.6 (0.4)
Amount Received in Interest	1.4 (0.3)	3.9 (0.7)
Amount Received in Wages	4.6 (0.3)	3.6 (0.3)
Property Value	2.7 (0.3)	3.6 (0.3)
Amount Paid for Water	2.5 (0.2)	3.5 (0.3)
Year built	2.9 (0.3)	3.0 (0.3)

Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

The ancestry question received the most requests for Help[11] (13.1 percent), and the second most requests relative to the number of people that saw the screen. We saw a similar result in the April analysis, suggesting this question continues to be problematic for respondents. However, this difficulty is likely related to the question concept itself and not the Internet version of the question. Therefore, to make the concept easier for respondents, we need to investigate whether this is something they do not know or that they do not understand. From there, we can research what aspect of the question is challenging for respondents and move to clarify that confusion within the question text.

[8] There was not a Help option for every question. Help was provided for the same questions as on the mail questionnaire. However, on the mail questionnaire Help is provided in a booklet that accompanies the questionnaire, not with the question as in the Internet instrument.

[9] The "Mobile home tax" and "Housing unit status (if vacant)" screens had question help rates of 12.4 and 3.0 respectively, but were only visited 92 and 67 times, respectively, as compared to more than 1,400 visits for the screens in Table 4. Therefore, additional research with a larger sample size is needed to determine whether a high percent of people actually needed help on these screens.

[10] The standard error was less than 0.05.

[11] The difference between the percent of clicks on Help on the Ancestry screen and the Pay for Electricity screen is statistically significant using a one-sided test at the $\alpha = 0.10$ level.

Unfortunately, for the other questions displayed in Table 4, we cannot say with any certainty why respondents are requesting Help or what type of information they are looking for. Moving forward into production, if the same screens continue to be associated with requests for Help, we may want to follow up with respondents to find out why they requested the help and whether they found the information they were looking for. Through the paradata, we will know which respondents requested Help on the same day that they made the request. This means that it would be possible to conduct a follow-up call immediately before the respondent forgets what he/she clicked and why.

In general, we think the help screens are worthwhile since most are used periodically by respondents, but not so frequently that we believe there is something problematic about the questions. There are a handful of questions that received very few or no Help requests. We should also continue monitoring these questions to determine whether the Help feature is even necessary for them.

5.1.4 Conclusions

We did not see anything surprising in the analysis of breakoffs, error messages, or use of Help. Respondents behaved as we expected based on the April paradata analysis. However, we confirmed that several screens identified in the April analysis should be monitored moving forward and also identified some areas of future research to minimize current issues.

First, we should research the impact of transitional screens on breakoffs. This can be through eliminating the screens or combining them with a screen with a question. As these screens were identified as having a higher proportion of breakoffs than other screens in both the April and November tests, there is reason to believe they will continue to account for the majority of breakoffs in production.

Second, as expected, there was a relatively high percent of errors rendered on the Place of Birth and Year Built questions. We know the majority of these error messages occur because respondents select the radio button but do not complete the write-in. As this question format continues to be problematic for some respondents, it may be worth experimenting with different formats to see if we can make the questions more user-friendly.

Finally, the Ancestry question continues to result in the most requests for help. It could be that respondents are just looking for what types of things count as ancestries. Alternatively, it could be the concept is difficult for respondents. If this is the case, we need to determine what part of the concept is challenging for participants and either change the question to make it more clear or ensure the help provides the information respondents need to answer the question.

5.2 Do respondents use the Help option and does it appear to be useful?

5.2.1 Do respondents use the Help option?
We have seen in past research that respondents do not often use available help features in Internet surveys (Conrad *et al.* 2003, Conrad *et al.* 2006, Lind *et al.* 2001). However, in the April Internet Test, 40 percent of Internet respondents selected Help at least once. Similarly, in the November Internet Test, 38.5 percent of respondents used a Help link at least one time.

The use of the Help feature was widespread across respondents and questions. Across the survey, the Help link was available on 107 screens and it was selected at least once on 97 percent of the questions for which it was available.

It is encouraging that we continue to see a high percent of participants using the help feature because this means they are able to locate the link and ideally receive the information they need to answer accurately. However, this does suggest that some respondents may be having trouble understanding and/or answering the questions. Therefore, it is important to ensure the help provided actually helps respondents answer the question or they may skip it, break off, or provide low quality data.

5.2.2 Is the Help option useful?
To determine how useful the Help was to respondents, we first looked at how many respondents used help multiple times. Additionally, we examined the percentage of cases in which a respondent selected an answer, clicked Help, and then changed their answer. Finally, we looked at the percentage of people who clicked help before selecting an answer and then provided a response after reading the help text. Unfortunately, due to the structure of the paradata, these percentages only include cases where the very next action was a field change.

In total, 2,038 respondents selected a help link. Fifty two percent of these respondents selected a link on more than one question and 31.7 percent clicked a link on more than two questions. While we cannot say so with certainty, respondents' willingness to continue to select help as they move through the instrument suggests they found useful information. Specifically, in their 2006 paper, Conrad and his colleagues found that respondents were more likely to continue using help if the information was useful.

Of all the instances that help was selected, 54.8 percent of the clicks were used to generate a response. In other words, they occurred prior to the respondent answering the question and resulted in selecting a response. In the April test, we also saw a higher percent of respondents using Help prior to selecting a response as compared to accessing Help after entering a response. These results suggest most respondents use the feature to help formulate an answer rather than to verify an answer. Only 5.1 percent of all the Help selections resulted in an answer change (after an initial answer had been selected). The low change rate could suggest that most respondents answered correctly the first time. However, it is also possible that the help text did not provide the information they needed to determine whether their original answer was correct. Regardless of whether

respondents are using the help to generate or verify an answer, it is important to make sure the help text provides the information respondents expect in a way they can easily understand.

5.3 Are there any problems with the re-entry authentication procedures?

To assess the re-entry authentication procedures, we analyzed the number of failed re-entries[12]. Once a respondent verifies his/her address, the respondent is provided with a four-digit PIN (Personal Identification Number). This PIN, along with the corresponding User ID, is used to log back into the instrument if the respondent leaves at any time. If the User ID and PIN do not match, the respondent will receive an error message.

Due to security regulations, we did not explicitly tell respondents to write down their PIN, but did tell them to "make note" of it. Therefore, there was initially concern that respondents would lose or forget their PIN and have difficulty re-entering the survey after a logout, timeout, or close-out. However, there were only 163 failed re-entries across all respondents. As a percentage of logins following a timeout or logout, 7.0 percent were invalid[13]. However, some of these respondents may have eventually re-entered the instrument. Additionally, respondents who had at least one invalid login averaged 1.6 invalid logins. This suggests that at least some respondents tried to access the survey multiple times before either giving up or entering. Although not many respondents had an invalid login, about 60 percent of those that did never gained access to the survey, which is similar to what we saw in the April analysis. Therefore, it may be worth experimenting with different authentication procedures. This may help reduce breakoffs by making it easier for people who try to return but cannot. However, the current procedures work well for most respondents, so any new procedures should be carefully tested to ensure they do not negatively impact the rest of respondents.

5.4 Is there additional information about the use of the instrument or survey features the paradata can help explain?

5.4.1 Devices used

The survey was accessed using a variety of devices. Overwhelmingly, the most popular was a standard personal computer. However, Table 5 provides a list of the variety of devices that were used to access the survey and how frequently each was used. We did not create a version of the survey specifically for mobile devices, so these users had to work with a survey designed to fit on a standard monitor.

[12] We were only able to measure attempted re-entries and not initial entries because in order to collect any data on respondents, they needed to log into the instrument at least once.

[13] We were only able to calculate the percent of invalid logins associated with timeouts or logouts, which only accounts for 49.7 percent of all of the invalid logins. The remainder is from respondents who closed the browser to exit the survey. We cannot calculate a similar rate for these cases due to complexities with working with the paradata.

Table 5. Devices Used to Access the Instrument

Devices	Frequency	Percentage (%)
Personal Computer	6,567[14]	95.5
iPad®	144	2.1
Windows® Tablet	102	1.5
Android® Phone	31	0.5
iPhone®	29	0.4
Blackberry®	2	0.0
Kindle Fire®	1	0.0
Total	6,876	100.0

Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

In total, 4.5 percent of devices used to access the survey were mobile. Specifically, 3.6 percent were tablets and 0.9 percent were mobile phones. We also searched for Android® tablets but were not able to find any.

Considering smartphones and tablets are becoming more popular and more devices are coming to the market, we believe more respondents are using mobile devices to complete the survey even though we cannot directly compare the data from the April test. The majority of people who used a mobile device to access the survey used a tablet. This trend should be investigated during production, and if we see increases in the use of mobile devices, we should consider developing a mobile-friendly version of the instrument.

5.4.2 Sessions to complete survey

While we allowed respondents to leave the survey and return using their assigned PIN, we prefer them to complete the survey in one session. This behavior reduces the chance they will forget to return or lose/forget their PIN. About 80 percent of respondents completed the survey in one session. Additionally, respondents averaged 1.3 sessions to complete the survey. These results are consistent with what we found in the April analysis (Horwitz *et al.*, 2012), suggesting we can expect to see the majority of respondents completing the survey in one session moving forward into production.

5.4.3 Average completion time

The time it takes a respondent to complete the Internet version of the ACS is dependent on household size and how many sessions it took them to complete the survey. Some individuals who accessed the survey logged out and then waited several days before returning to the instrument, likely due to receiving another mailing. Response times that include individuals that completed the survey in multiple sessions have large outliers and high standard deviations. Therefore, this analysis only includes cases that completed the survey in one session because these results are more meaningful.

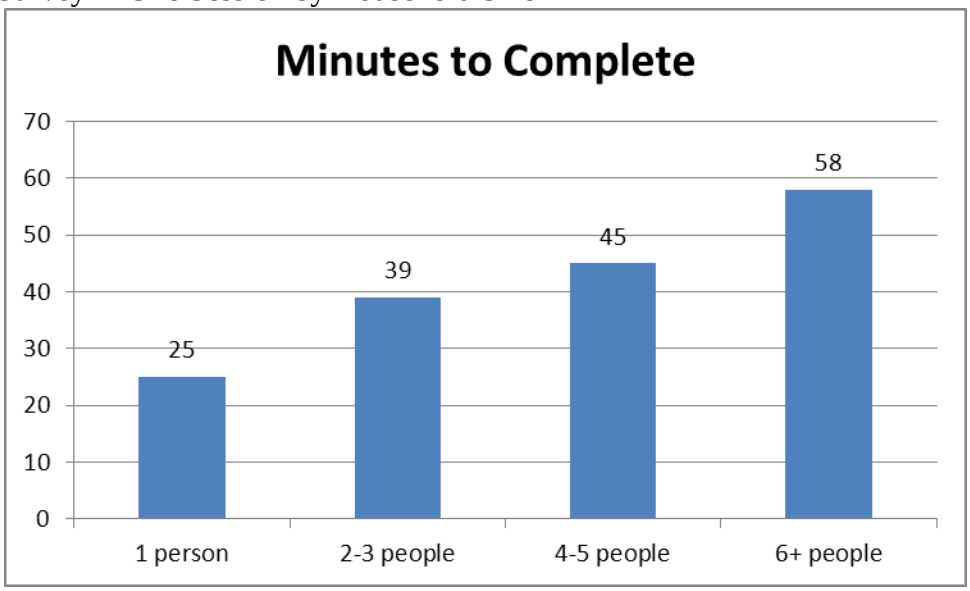
In this analysis, we calculated how long it took the average respondent to complete the

[14] The total number of devices used is greater than the 5,291 households that accessed the instrument because it reflects every time the instrument was accessed.

survey and also how long it took when controlling for household size. The average completion time was calculated by subtracting the time a respondent entered the "Address Verification" screen from the time he or she entered the "Presummary" screen. Included in this calculation are the individuals who viewed all of the applicable ACS questions in one session. For all respondents who completed the survey in one session, the mean completion time was 37 minutes. This estimate is about what we expected based on the time estimate displayed on the PIN screen (38 minutes).

We also looked at completion time by household size for respondents that completed the survey in one session. Figure 3 provides the median completion time for the total survey by household size.

Figure 3. Median Completion Time (in minutes) for Households that Completed the Survey in One Session by Household Size



Source: U.S. Census Bureau, 2011 ACS Follow-up Internet Test, November to December 2011

Figure 3 shows that completion time increases as household size increases, which is what we expect because larger households have more questions to answer than smaller households. Additionally, we saw a similar pattern in the April test.

6. SUMMARY

This paradata analysis was intended to help us understand how respondents interact with the Internet instrument, to identify any potential problems with specific questions or with the instrument as a whole, and to identify any differences from what we saw in the April test. To do this, we looked at breakoff rates, error messages, use of the help feature, devices, authentication, and completion behavior. The overall finding from this research is that the Internet instrument is working as we expected it would. Respondents do not seem to be having too much difficulty accessing the instrument and they are able to navigate through the instrument and use the available features. The majority complete the survey in one session and are completing the survey in the amount of time expected.

Similar to the April findings, the Internet instrument continues to work well. We did identify some areas where we may be able to improve the instrument, such as reducing breakoffs on transitional screens and reducing error messages by changing the format of the Place of Birth and Year Built questions. While we did not uncover any new issues as compared to the April Internet Test paradata analysis, we have identified some areas that we should continue to monitor into production, such as the devices used to access the survey, the number of sessions to complete the survey, completion times, and where respondents are requesting help.

Acknowledgements

We would like to thank the following Census Bureau staff for their valuable contributions and assistance to the development and analysis of this project: Brenna Matthews, Andrew Roberts, Brian Wilson, Tony Tersine, John Studds, Chris Butler, Joe Mistichelli, Brian Ridgeway, Michael Coan, Fred Meier, and Samantha Barron.

References

Conrad, F., Couper, M., Tourangeau, R. (2003). Interactive Features in Web Surveys. Paper presented at the Joint Statistical Meetings of the American Statistical Association. San Francisco, CA.

Conrad, F., Couper, M., Tourangeau, R., Peytchev, A. (2006). Use and Non-use of Clarification Features in Web Surveys." *Journal of Official Statistics* 22(2): 245-269

Couper, M.P. (1998). Measuring Survey Quality in a CASIC Environment. *Proceedings of the Section on survey Research Methods of the American Statistical Association*. Dallas, TX.

Galesic, M., and Bosnjak, M. "Effects of questionnaire length on participation and indicators of response quality in a web survey." *Public Opinion Quarterly* 73.2 (2009): 349-360.

Horwitz, R., Tancreto, J.G., Zelenak, M.F., Davis, M.C. (2012). Use of Paradata to Assess the Quality and Functionality of the American Community Survey Internet Instrument.
http://www.census.gov/acs/www/Downloads/library/2012/2012_Horwitz_01.pdf

Lind, L., Schober, M., Conrad, F. (2001). Clarifying Question Meaning in a Web-Based Survey. *Proceedings of the Annual Meeting of the American Statistical Association*. August 5-9.

Matthews, B., Davis, M.C., Tancreto, T.G, Zelenak, M.F., Ruiter, M. (2012). 2011 American Community Survey Internet Tests: Results from the Second Test in November 2011. http://www.census.gov/acs/www/Downloads/library/2012/2012_Matthews_01.pdf

Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly* 73(1): 74-97.

Tancreto, J.G., Davis, M.C., Zelenak, M.F. (2012a). Design of the American Community Survey Internet Instrument.
http://www.census.gov/acs/www/Downloads/library/2012/2012_Tancreto_02.pdf

Tancreto, J.G., Zelenak, M.F., Davis, M., Ruiter, M., Matthews, B. (2012b). 2011 American Community Survey Internet Tests: Results from the First Test in April 2011.
http://www.census.gov/acs/www/Downloads/library/2012/2012_Tancreto_01.pdf

Appendix: Paradata Collected

Category		Paradata
Survey Access	-	Number of users that access the survey via a mobile browser
Login/Authentication	-	Number of times an incorrect User ID is entered
	-	Number of times an incorrect User ID/PIN combination is entered per User ID
	-	Number of times a respondent is locked out of the survey because of an incorrect User ID/PIN combination
	-	Number of times a respondent tried to come back into a survey that was already completed and submitted
Number of Sessions	-	Number of Sessions per User ID
	-	Number of respondents that completed the survey in multiple sessions
	-	Average number of sessions needed to complete the survey
	-	Time between sessions for a respondent
Answer Changes	-	On which screen(s) does a respondent change an answer (radio button)
	-	Number of times a respondent changes an answer on a particular screen
	-	What screens a respondent visits more than once
Error Messages	-	On which screen(s) does a respondent get an error message
	-	Number of times a respondent gets an error message
	-	Number of times a respondent selects an answer after receiving an error message
	-	Aggregate number of times an error message is rendered on a particular question screen
Help	-	For what questions does a respondent click "Help"
	-	Number of questions for which a respondent clicks Help but does not give an answer
	-	Number of times a respondent clicks "Help"
	-	Number of times "Help" is clicked for a particular question