Consumer Expenditures Proof of Concept: A Joining of Quarterly and Diary

Michael K. Mangiapane – U.S. Census Bureau

1. Introduction

The Consumer Expenditures Survey project (CE) is a long-standing survey conducted by the U.S. Census Bureau under the direction of the U.S. Bureau of Labor Statistics (BLS). It originally started out in the late 1800s as a survey that was conducted approximately every ten years until 1980. The data obtained from the survey is used to update an important economic indicator known as the Consumer Price Index (CPI) which measures price inflation in the U.S. economy. Due to a need for more frequent data about consumer spending patterns and faster updates of the CPI, BLS changed CE to be a continuously running survey made up of two instruments. These instruments became the Consumer Expenditures Quarterly (CEQ) and the Consumer Expenditures Diary (CED). While these instruments have been serving the data collection needs of BLS, they are always looking for new and improved ways to collect expenditure data. In this paper, we will discuss what could become the next generation of CE instruments and the steps we are taking to achieve BLS's needs.

2. Design and History of CE Instruments

CEQ and CED interview separate samples of households in the United States. A household that falls into either sample is interviewed as a single group of people, also known as a Consumer Unit (CU). If there is anyone living at the household that pays their own living expenses, they are interviewed as a separate CU and added to the sample. A household that falls into the CEQ sample is interviewed every three months for a total of five interviews (this was revised in February 2015 to four interviews). Each interview consists of collecting data about the expenditures that the CU had made for the past three months. This instrument is designed to capture data about large expenditures that a respondent would recall easily, such as the cost of rent or a mortgage, spending on furniture purchases, vacations, car payments, and educational expenses. A household that is a part of CED is visited two or three times during a three-week period. A brief interview is conducted with the CU before they are given a paper diary to record all of their expenditures for one week. The first diary is picked up and a second one is given to the CU to record another week of expenditures, for a total of two weeks of recorded expenditure data. The idea behind CED is to gather data on expenditures that a CU may not recall in a regular interview (e.g., CEQ) such as a visit to a convenience store or a single purchase at a fast-food restaurant. The only expenditures that are not included in either survey are business-related expenditures or expenditures where the CU was reimbursed.

Both CEQ and CED started out as PAPI surveys. Since there are many different types of expenditures that a CU can make, CEQ became a very large survey. Before its conversion to CAPI, a CEQ form would have a minimum of 76 double-sided ledger-sized (11 inch by 17 inch) double-sided pages to be filled out. If there was more than one person in the CU or they owned more than three properties, extra pages would have to be added to the appropriate sections. CED's form was made up of 10 double-sided ledger-sized pages of questions about the household along with a section to record any diary entries that may have been missed by the CU but recalled when the Field Representative (FR) returned to collect each diary. The diaries themselves were 26 double-sided letter-sized pages. As one can imagine, with the

costs of printing and processing each of these instruments, the CE surveys were excellent candidates to converted to CAPI.

In 2000, an initiative was undertaken by the Census Bureau and BLS to convert CEQ and CED from PAPI instruments to CAPI instruments using Blaise to implement these instruments. The use of Blaise was significant because CEQ and CED were two of the first instruments to be developed in Blaise at the Census Bureau. After a period of development and testing, CEQ went into production in the field in April 2003 and CED went into production in January 2004.

2.1 Current CE Instruments

Currently CEQ is running in Blaise 4.8.4.1861 and it contains twenty-two different expenditure sections. Additionally, there are sections in the instrument to collect the demographics of the CU and the characteristics of the household. After the demographics are collected, the FR asks questions about the CU's expenditures for the past three months, including related follow-up questions like if purchased items were bought for the CU or someone outside the CU and if taxes were included on the reported expenditures. Sections that ask about income are asked only during the second through fourth interviews, and a section about financial assets is only asked on the fourth interview. There is also a debriefing section in the back of the instrument that collects information about any records that were used during the interview and any aids that may have been used, such as an information booklet or a home file. All available sections in the instrument for the interview are accessible on the main path or they can be accessed via parallel tabs, giving the interviewer the ability to jump to specific sections during the interview. CEQ has kept a similar design for CAPI as it had for the PAPI version, but with the flexibility to collect more expenditure data during an interview. However, the CEQ instrument does make for long interviews with respondents. The average interview requires 1 – 1.5 hours to complete. This may have an effect on response rates for the survey. Additionally, moving to CAPI has allowed the sample size to increase. Since 2005, approximately 5,000 households in the U.S. are interviewed every month for CEQ.

CED is running in Blaise 4.8.4.1861and contains a nearly identical demographics section as CEQ. Once that information is collected, the CU is asked a small subset of questions such as if they own their home, what they spend on groceries, and how many cars they own. There is also a section that asks about each CU member's employment status and income information, just like in CEQ. That section is optional for the first two visits, but it must be completed during the final visit. The FR leaves a paper diary for the CU to record their expenditures for a 1-week period. After the initial week, the FR picks up the first diary and leaves a second diary for the CU to fill out, and the FR asks some follow-up questions about the first diary to prompt the CU to recall any expenditures they may have potentially forgotten to record. For the final visit the FR will pick up the second diary and collect income information from the CU if that has not been done already. Approximately 2,000 households are interviewed for CED every month.

While CED is a smaller instrument than CEQ, it does have some unique requirements that required being creative when we coded the instrument. The instrument is date-sensitive in that the FRs are given a window in which they are to make each of their visits. The first window is provided from an input file and the FR has six days to make initial contact with the household and leave a diary for the CU. After the diary is placed, the instrument calculates a date one week in the future and gives the FR and the respondent a seven-day window to pick up the first diary and leave the second one. After the second visit, the instrument calculates another date one week in the future and creates a seven-day window for

the FR to make their final visit. There are also times where it may not be feasible for an FR to make three visits to the CU because of travel distance, workload, or because the CU knows they will be unavailable during one of the visit windows. The FR is allowed to do what is called a "double-placement" in that they place both diaries with the CU on the first visit. In this scenario the instrument calculates a date that is fourteen days in the future and creates a seven-day window for the FR to return. The instrument also has to skip what is considered to be the "normal" second visit and combine all follow-up questions to be asked on the final visit. All of these date calculations must be passed back to the FR's case management software so that they have a central place where they can reference when and where they need to make their next visits.

3. Project Gemini and CE Proof of Concept

The basic methodology of CEQ and CED has remained the same since 1980, with the biggest change being the conversion from PAPI to CAPI but still using similar sections and questions. With rising costs to conduct a survey and a desire to reduce respondent burden, lower measurement error rates, and leverage the latest technologies that are available, BLS is exploring ways to update the CE instruments so interviews are conducted more efficiently without a negative impact on data quality. They have decided that a major re-design of the CE surveys is in order. They have an initiative known as the Gemini Project to research proposed changes to CE and develop what will be the next generation of the CE instruments. At this stage in the Gemini Project, BLS is working with the Census Bureau to develop a new instrument that combines CEQ and CED into one single instrument that is split into two interviews to collect CEQ information, a Records interview, and a Recall interview. Since both CE instruments are already in Blaise, theoretically they can be combined together using common code blocks, but there will still need to be adjustments made to this combined instrument. There is also a web-based instrument/app that is being incorporated into this survey, which will be used to collect diary data for individuals in the CU (rather than a single diary for the whole CU) over a one-week period. The idea is that respondents should access a mobile site or an app on their phone or tablet at the time they are making their purchase during the diary period, in order to obtain more detailed and accurate diary data. Since this new instrument for the Gemini Project is in the early stages, it is known as the CE Proof of Concept, CE POC, or simply POC.

CEQ and CED are very different in their approaches to collect expenditures, but they do have some overlap in the information they collect. Both instruments use similar household demographic sections, and they ask the respondents about their income, if they have a mortgage, how many cars they own, and their spending on groceries. However, CEQ asks more detailed questions about expenditures. In considering the design of the Proof of Concept instrument, we decided to use the current CEQ instrument as the baseline and add the appropriate CED sections as needed. The sponsors reduced the number of CEQ sections from 22 to 16 and split the collection of the sections into two separate interviews. The first set of CEQ sections make up the Recall interview since they ask about expenditures that a respondent is likely to remember, such as what they spent on a vacation or on clothing. The second set of interview sections are the Records interview because the CU is likely to have records of those expenditures, such as mortgage or rent payments. What the POC instrument did get from CED were the sections that related to placing and picking up the diary, including the "recall" table to add additional expenditures that the CU did not enter in their diary but remembered during the second visit. Besides the CED sections, other code would need to be added to make the POC fit the requirements that the sponsors wanted.

4. Outcome Codes (Control Codes)

All CAPI surveys conducted by the U.S. Census Bureau have a set of outcome codes (control codes) that are assigned to each case. These codes are status codes that tell our Laptop Case Management (LCM) software to perform an action on the case and allow Headquarters staff to see the status of the case. The codes are also an easy way to reference what happened with a case, such as the reason a case was a non-interview. CEQ is very straightforward with its outcome codes in that they are set as an overall code for the entire case. However, CED not only has an outcome code for the entire case, there are separate outcome codes set for placing the diaries or picking up the diaries. This is to account for all potential possibilities that can occur during each visit, such as a respondent who filled out a diary one week but did not fill it out another week, a diary being refused, a respondent who is out of town during the first week, or both diaries being completed. Since POC involves diaries, we decided to use the same LCM interface that CED uses. The POC needed to set the same diary placement and pickup codes along with overall outcome codes for a case. There were also a few new outcome codes added to track whether or not the Recall or Records interviews were completed.

The sponsors had to consider all possible results for a case, and it meant making sure the instrument could handle each result and set the appropriate outcome codes. This was important because some outcome codes determine the flow of the interview. For example, if a non-interview outcome code is set, the instrument will skip over all middle sections and go straight to the back. There is also an FR debriefing section that only comes on-path when the Recall Interview is complete and the diary has been placed with the CU (or it was refused, or the interview was completed after the interview date window).

In the following example, the case on its second visit (VISITNUM) and the diary pickup (Outcome_Pickup) and Recall Interview (Outcome_Recall) were completed. However, the Records Interview is not complete (Outcome_Records = 202) so the case is given an overall outcome of 383. Each Recall or Records section has a flag that tells the instrument if that section is complete during the current interview. Once all of the sections are completed for their interview, the appropriate outcome code is set and that code will change the overall case outcome code.

```
IF SectComp[2] = YES AND SectComp[3] = YES AND SectComp[4] = YES AND SectComp[10] =
YES AND SectComp[11] = YES AND SectComp[13] = YES AND SectComp[14] = YES AND
SectComp[15] = YES AND SectComp[21] = YES AND SectComp[22] = YES THEN
     Outcome_Record := '201'
ELSEIF SectComp[2] = YES OR SectComp[3] = YES OR SectComp[4] = YES OR SectComp[10] =
YES OR SectComp[11] = YES OR SectComp[13] = YES OR SectComp[14] = YES OR SectComp[15]
= YES OR SectComp[21] = YES OR SECTCOMP[22] = YES then
     Outcome_Record := '202'
ELSE
     Outcome_Record := EMPTY
ENDIF

IF VISITNUM = 2 AND  Outcome_Recall = '201' AND (
(Outcome_Pickup <> '201' AND Outcome_pickup <> '325' AND Outcome_pickup <> '326' AND
Outcome_Record <> '201') OR
((Outcome_Pickup =  '201' OR Outcome_pickup =  '325' OR Outcome_pickup = '326' ) AND
Outcome_Record <> '201') OR
(Outcome_Pickup <> '201' AND Outcome_pickup <> '325' AND Outcome_pickup <> '326' AND
Outcome_Record =  '201'))AND Admin.ReplaceSpawn <> Yes THEN
    Outcome := '383'   {Visit 2 follow-up (pickup or records incomplete)}
ENDIF
```
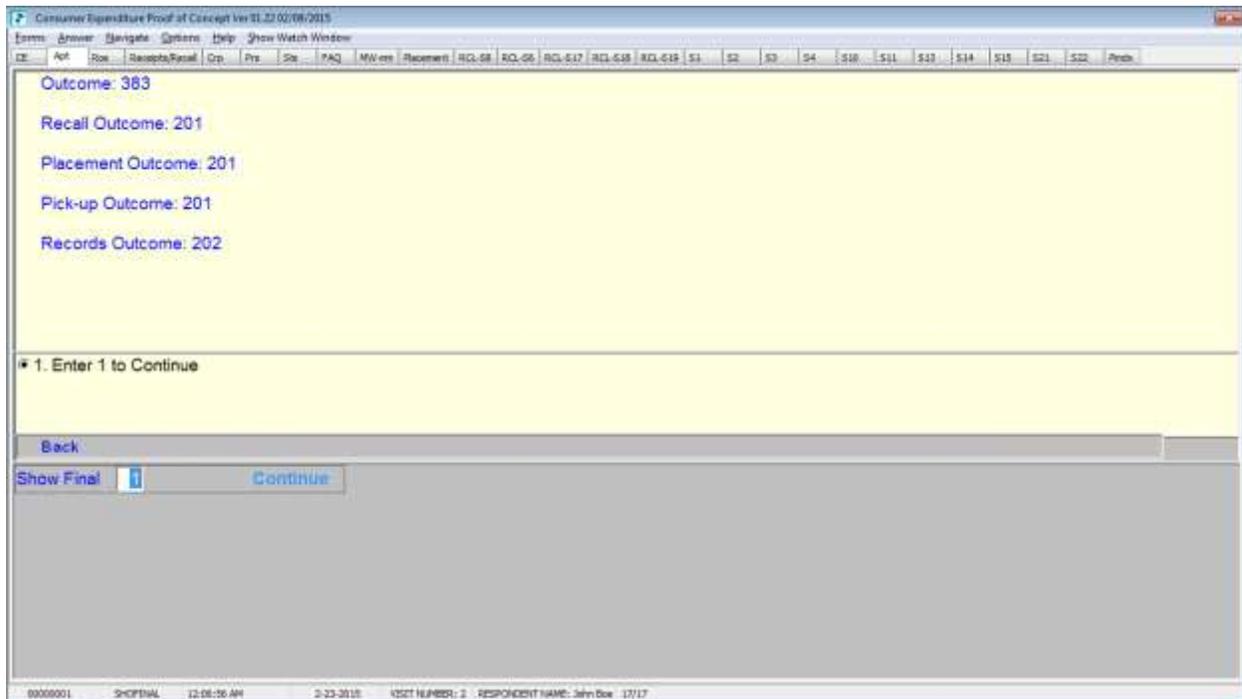
Fig. 1 – Display of all outcome codes that are set in the POC. "Records Outcome" is not a completed outcome code therefore the current overall case outcome code is set to 383, "Follow-up Required (Visit 2)"

4.1 Outcome Codes Everywhere

With CE being one of the first surveys written in Blaise, it meant that there was a period of adjusting our thinking about how to program instruments in Blaise vs the previous authoring language, CASES. One example of this is how the outcome codes were set. The normal approach for setting outcome codes in CASES was to set the outcome code immediately after reaching the condition in the survey that satisfies the outcome. So setting each outcome code was placed in-line in the instrument code during the execution of the survey, meaning outcome codes were being set throughout various sections. With Blaise's rules re-execution engine, that isn't necessary since the rules are checked all the time. Therefore, an instrument can have a separate block or procedure that will run at the end of the rules in the datamodel. However, the original implementation of CEQ and CED used some CASES style methodology in there. While the setting of outcome codes worked, it also presented difficulties when outcome codes needed to be changed since multiple sections of code had to be checked and verified before the change could be implemented. Despite the maintenance challenge, the outcome code assignments were left in place because they were working and there was less of a desire to risk causing a problem in production.

4.2 Outcome Code Procedure

Since we were making some major changes to CEQ's code to accommodate CED and we had to add the CED placement and pickup outcome codes, we decided that we would move setting all of the outcome codes in the POC instrument into a procedure. This meant finding every single outcome code assignment and moving them into a single procedure that would be called in the back of the instrument. Within the new outcome code procedure, we would write the appropriate universe statement for each outcome code assignment and thoroughly test the procedure to make sure all of the expected outcomes were reached. This included not only the overall case outcome codes, but the diary placement and pickup outcome codes

as well. Since CE has a well-established set of testing scenarios and we implemented this change early in the project, we were able to conduct a very thorough internal test before the instrument went out for system interface testing. We immediately saw the benefit of using the single procedure as we were able to easily test and examine why an outcome code was or was not what we expected. It was also beneficial to have the expected conditions for assigning that code right there in the universe statement. We did not have to search our code to determine if an outcome code assignment was buried inside a nested universe in some other section of code. This approach will make future changes to the outcome code easier as only one section of instrument code has to be tracked and maintained instead of several sections of code.

5. Recall and Records Interview Sections

With the CEQ sections being split into the Recall (Visit 1) and Records (Visit 2) interviews, that meant shuffling the sections so that they are asked in the correct order. For example, the sponsors decided to ask Sections 8, 6, 17, 18, and 19 (in that order) for the Recall interview and the rest of the sections would be asked in the Records interview. Since each CEQ section is an individual block in Blaise that is called from the main datamodel, setting up the Recall and Records sections were a matter of moving the calls to the sections to the appropriate places in the rules of the datamodel. Additionally we added a universe to check and see if the value of VISITNUM was 1 or 2 so that the appropriate sections would be put on path.

The sections are declared as parallel blocks so an FR can complete the sections in a different order if they want to. Having parallel blocks was also helpful when a new requirement was added to have the instrument not ask the Recall sections again during the Records interview, but still allow the FR to have access to the Recall sections. If for some reason the CU forgot something in the previous visit, the FR can still access those sections via parallel tabs and add that information. However, the questions are not on the main instrument path during the Records interview so the FR does not have to press Enter multiple times or the End key to get past those sections. Blaise has been flexible in allowing the instrument to keep the proper sections on path without losing data or making the FRs repeat sections that are not required for that interview.
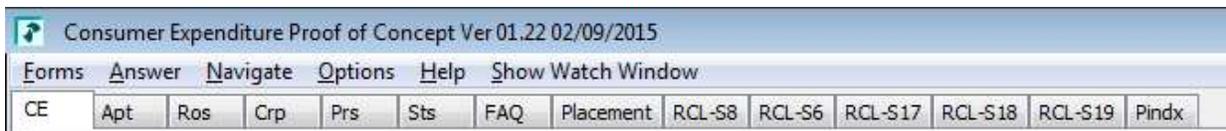

Fig. 2 – All available sections in the Recall Interview

Fig. 3 – All available sections in the Records Interview

Many more sections are put on path in the Records interview. Besides the Records sections, there is also a section to show the list of usernames assigned to the members of the CU if they opted for a web diary. The Receipts/Recall section is also on path for the FR to enter any expenditures that were not recorded in a diary but were recalled by the CU during the interview. The Recall sections are still available in case any additional expenditures need to be added during the Records interview.

6. Tracking Diaries

Another challenge in programming the POC instrument was how to properly track the diaries. In a regular CED interview, the FR leaves a paper diary that is filled out by the CU. When the FR returns and picks up

the diary, they send it back to their regional office (RO). The RO takes that diary, assigns it a barcode, and it is shipped to the National Processing Center (NPC) in Jeffersonville, Indiana for scanning and processing. With the Gemini Project, there is a requirement to move away from using paper diaries as the means of collecting expenditures for CED. In 2013 a test was done with a sub-sample of CED respondents that gave them the option to use a paper diary or to fill out the diary on the Internet using a provided username and password. For that test, each CAPI case was pre-assigned a username and password that the FR would give to the respondent if they opted to fill out a "Web Diary." In 2014, a different test was conducted on a sub-sample of CED respondents, known as the "Individual Diary" test. Each eligible member of the CU was given the option to fill out a Web Diary using their computer or a mobile device (an eligible member of the CU is a member over the age of 16 who is currently living at the household). Instead of one diary for the entire CU, there was a diary for each eligible individual of the CU who opted in to filling out the diary. For that test, no paper diaries were used for the sub-sample. If they wanted a paper diary, they were screened out.

The methodologies from both of these tests are being combined for the diary component of the POC. The goal of this new Web/App Diary is to allow respondents to record their expenditures at the time of purchase using an app on their phone or tablet in order to capture more accurate and complete diary data. After the Recall Interview is completed, each eligible member of the CU is asked if they regularly access the Internet. If they do, they are given the option to fill out a Web/App Diary. A member who opts for the Web/App Diary is given a pre-assigned username and password so they can access the diary throughout the collection period. Due to security requirements, the password is required to be long and complex. This raises usability concerns as the respondent is required to key in this potentially challenging password every time they open the app. However, a member of the CU may instead choose to fill out a paper diary or they may decide they do not want to participate in recording their expenditures. The instrument will capture how many paper diaries were given out so that the RO and NPC know how many paper diaries to expect for processing.

6.1 Web/App Diary Logins

Being able to provide diaries to individual members of the CU and have them fill them out on paper or on the web introduces a number of possibilities and requirements that must be considered for the POC instrument and for survey procedures. A big requirement to consider is security and being able to have enough web diaries for all members of the CU. The POC's household roster allows for up to thirty members in a household, and theoretically they could all be part of the same CU. The Centurion software that is used for the Web/App Diary does not allow for custom accounts to be created. Even if it did, there would need to be a way to connect that web diary with that POC case. It is also possible that an FR may be conducting the interview in an area where they may not be able to access the Internet to allow for account creation. Even if there is Internet access available, there is a question of making sure the connections and any data being transmitted is secure.

In considering some of these concerns, the sponsors decided that the POC instrument would display the web diary accounts needed based on the household roster, and that the accounts themselves would be pre-created in Centurion. Since each POC case is a single Blaise database, we stored all of the potential usernames and passwords needed into one block. Once the FRs collect the household roster information, the instrument determines the appropriate CUs. If a CU member opts for the web diary, the instrument will display a username and password that is given to the member to access their web diary. For security

purposes, after the web diary username and password is provided to the CU member(s) on the first visit, the FR will not be able to access information about the passwords from the instrument. Instead, we added a screen in the instrument to let the FR see a web diary username only, in case a CU member forgets that and calls the FR. There is a help desk that a CU member can call if they need to have their password reset. Otherwise, once the username and password is given to the CU member, accessing the diary and filling it out is completely their responsibility.
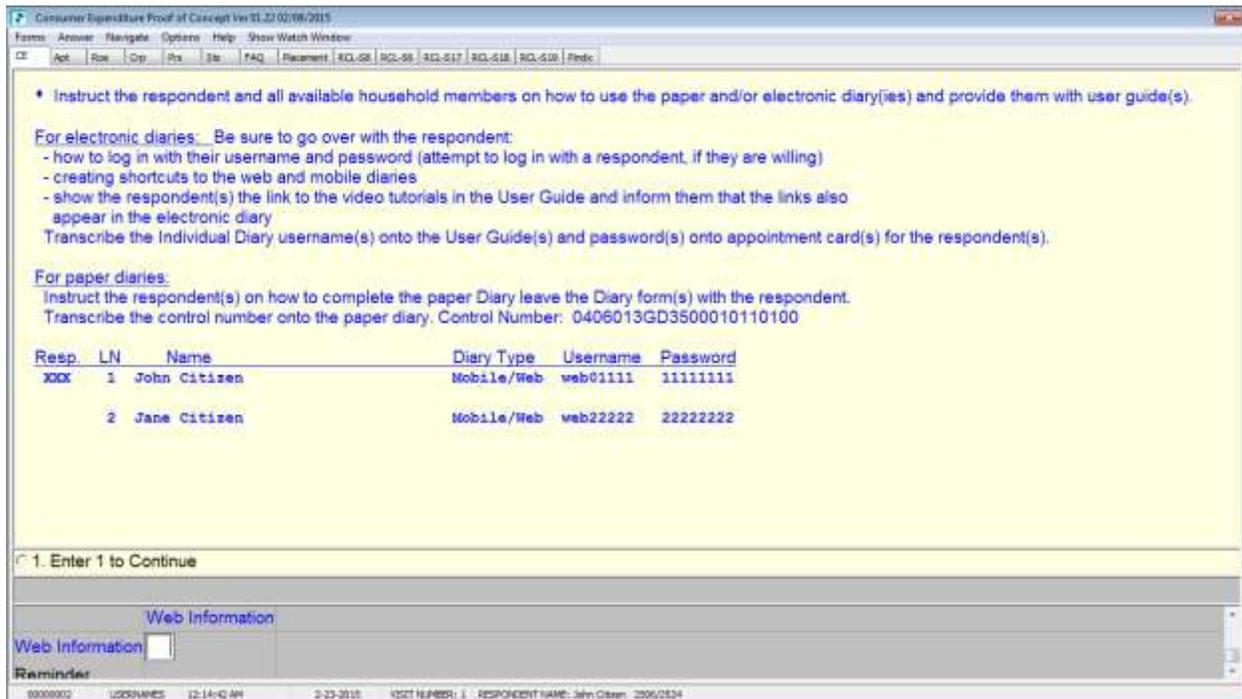
Fig. 4 – Screen where the FR can provide login information to the CU members at the time of diary placement

6.1.1 Spawned CUs

As we were programming the POC instrument, we realized there was a possible situation that could arise with a case and its handling of the usernames and passwords. As mentioned in the design of the original CE instruments, a CU is a group of people living in the household and typically there is one CU per household. However, there are households where you might have multiple CUs, due to living situations such as roommates or multiple families who each contribute their own money to household expenses and make expenditures exclusively for themselves. Both CED and CEQ have sections that ask if certain members of a household (such as non-relatives) pay for their own expenses with their own money. If the instrument determines that a household member is a part of a different CU, the instrument and the LCM software work together to create a separate case for the FR to interview the other CU. To accomplish this task, we run a Manipula script after the FR exits the case. The script will generate a new Blaise database and populate it with information from the original case. LCM handles adding the newly spawned case to their database so that it is properly added to the FR's assignments and is transmittable back to headquarters when it is completed.

Since the instrument can spawn new cases to accommodate multiple CUs, the question became how to handle the new case so that a spawned CU would also have access to a Web Diary to fill out their information. Since the Web Diary accounts have already been made, there currently is not a way to create

them on the fly. If that was possible, we could use Blaise to calculate a new username and password for the FR to give to the CU members who opt for a Web Diary. Since that was not an option, we decided to copy over some of the unused usernames and passwords from the parent case to the spawned case for the members of the new CU so that they would have access to the Web Diary. Since each username and password is connected to one line of the roster, it is a matter of copying over the appropriate username and password via Manipula when the spawn case is created. However, this approach means that any household members added to the roster in the spawned case will not have a username and password available for them. This should be a very rare occurrence since their name and status as a household member should have already been collected in the parent case.

As an example of how we made this work, we could have two couples living in the household that each pay their own expenses. Since that would create two CU's, Person 1 and Person 2 on the roster are their own CU and stay with the parent case. Person 3 and Person 4 will be interviewed in the spawned case. The username and password that would be available for Person 3 and Person 4 in the parent case is copied over into the spawned case via Manipula. We have found in our own systems testing that this does work without an issue.

7. Using the Web/App Diary

When a user logs into the Web/App Diary for the first time, they will enter the date they were told to start recording their expenditures. After that they are taken to a screen to enter their expenditures.

7.1 Internet Version of the Web/App Diary

Fig. 5 – The main screen of the Web/App Diary

Selecting a category will bring up additional information to be entered about their expenditure.

Fig. 6 – Follow up information about "Food and Drink Away from Home" purchases.

Fig. 7 – Follow up information about "Clothing, Shoes, Jewelry, and Accessories" purchases.

As expenditures are added, they are displayed under the summary of expenses. The user is also able to edit or delete expenditures.

Fig. 8 – Editing a Web/App diary entry.

7.2 Mobile Device Version of the Web/App Diary

Fig. 9 – App Main Screen   Fig. 10 – Adding a new expense in the App   Fig. 11 – After adding an expense
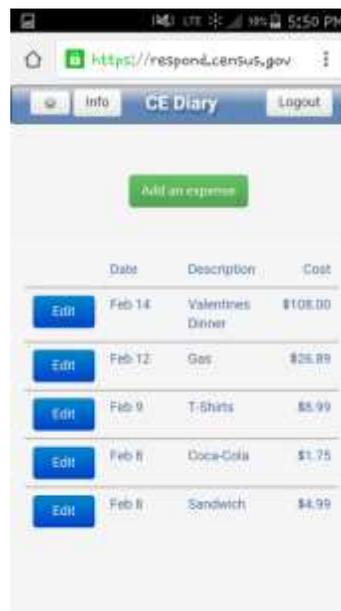
8. Future Developments

The long-term goal of the Gemini Project is to create a system that more efficiently and accurately collects the data required for the CE survey. To accomplish this goal, the Census Bureau hopes to use a combination of data collection methods that seamlessly interfaces with their control and tracking systems so that the staff at Headquarters can examine the status of each case and its individual diaries in real-time. This summer, we will be fully testing the POC instrument using the combination of the Blaise instrument to conduct a majority of the CE interview and the Web/App Diary (developed in Centurion) to collect the individual diary information. The results of the test and the lessons learned from programming this instrument will be analyzed and used to enhance the overall design and implementation of the redesigned CE survey.

9. Conclusion

For a large and complex survey like CEQ or a smaller and date-sensitive instrument like CED, Blaise has handled the requirements from BLS and from the Census Bureau with success in collecting crucial economic data. However, Blaise does not have to do everything all by itself. By leveraging other available technologies that provide other approaches, we can help streamline the data collection process. If we can use them to interact with the Blaise instruments and data, it opens up further opportunities to tap into the power of Blaise and what it can do with our data. The POC has already provided new ideas for us to consider applying to other surveys in the future.

10.   Acknowledgements

The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau.