

# Using Technology to Conduct High-Volume Cognitive Interviewing: Lessons Learned

Melissa A. Cidade

Kristin J. Stettler

Kelsey J. Drotning



Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. The U.S. Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY22-ESMD002-015)

1

**8,792 PET STORES\***  
**+**  
**\$18.9 BILLION IN DOG AND CAT FOOD PRODUCTION\***  
**=**  
**39,593 BUSY CARPET CLEANING PROFESSIONALS\***

**THE ECONOMIC CENSUS.  
SEE HOW IT ALL ADDS UP.**

Learn more at [census.gov/EconomicCensus](https://census.gov/EconomicCensus)  
\*Source: U.S. Census Bureau—2012 Economic Census of the United States

United States  
**Census**  
Bureau

2

The research we are talking about today is in the service of the 2022 Economic Census. The Economic Census is a mandatory survey conducted every five years – a quinquennial data collection – in years ending in 2 and 7. The survey collects data electronically from nearly 4 million businesses (including large, medium and small companies representing all U.S. locations and industries) on a range of operational and performance questions at the establishment level. Data from the survey are used as the official five-year measure of American business and the economy. The data from the Economic Census are used as the foundation of the Gross Domestic Product and other leading indicators, as well as updating the Business Register, the most current and comprehensive database of U.S. business establishments and companies for statistical programmatic use.

Throughout this presentation, I'll refer to the 2022 Economic Census as EC22 for short.

## Components of High-Volume Testing

- Recruitment
- Test instrument
- Interviewing protocol
- In-field support
- Progress tracking
- Analysis

### By the Numbers:

- 230 cognitive interviews
- 19 topics
- 65 questions
- 3 rounds of interviewing
- 6 interviewers
- 4 months (August – November 2021)



3

Every five years, we re-examine the content of the Economic Census to ensure that we are collecting high quality data on topics pertinent to the American economy. Due to the size and scope of the Economic Census, testing new content is one of the largest recurring projects our staff works on. This year, we were asked to test 65 questions spanning 19 distinct topics within the survey. I will not be dwelling on the content of the testing, but rather, this presentation will focus on how we conducted 230 interviews across all sectors of the economy in a staggeringly short 4 month period. On a typical project, we average about 25 interviews; for this project, we completed more than nine times that number.

Throughout this presentation, I will be sharing some of the tips and tricks we have learned using various computer programs. Note that this is not an endorsement of any of the programs mentioned during this talk, but rather serves as a framework of how to conduct high volume cognitive testing. I'll talk through recruitment, the development of a test instrument, the interview protocol, in-field support, progress tracking, and analysis during this talk.

## Recruitment – Who are we targeting?

- Very specific targets
  - Size
  - Location
  - Kinds of activities
- High-volume
- Rolling basis
  - Flexibility during interviewing
  - Met specific recruiting targets weekly

### Solution:

Training to pull cases directly from the Business Register.



4

The first issue we needed to tackle when thinking through this project is how we were going to get recruitment lists. Working with key stakeholders within the Census Bureau, we identified the various types of businesses we wanted to ensure were included in testing. This was sometimes very broad – say, manufacturing companies, companies from a specific region or state, or even some questions that were sector-neutral and needed to be tested with a wide berth of companies. For other questions, however, the target population is very narrow – for example, we tested a few questions specifically designed for space vehicle manufacturing companies, a new and emerging industry. At the same time, we needed a lot of companies, and we needed the lists on a rolling basis. That is, we needed the recruitment sample to be flexible as we completed interviews so that we could continue to hit the very broad and very specific needs of this project.

The solution we came up with was both radical and simple: training on how to pull our own data. Traditionally, the cognitive testing team works with the survey specific staffs to identify cases for testing. However, since the EC spans sectors of the economy, and the need for flexibility throughout the testing period, we ultimately worked with mathematical statisticians to gain training and access to the Business Register to pull our own recruitment sample as the project needed. This included several hour-long walkthroughs of accessing the data, understanding the variables, and selecting on the correct criteria. While the learning curve was steep, it ultimately allowed us to be as flexible as the project needed, and built out a new capability: pulling our own recruitment files.

## Recruitment – How will we recruit?

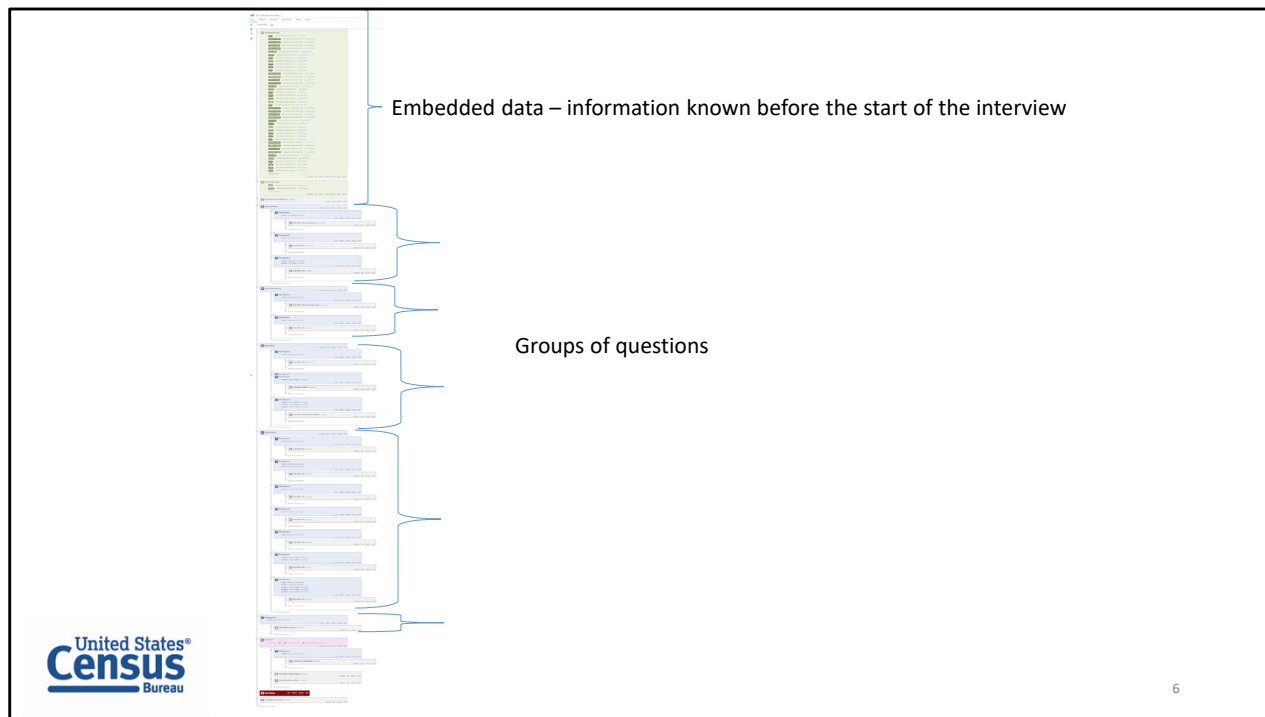
- Self-scheduler – respondents pick the interview day and time
- Weekly deployment – tailored to interviewer availability
- Daily tracking – to meet recruitment goals

A screenshot of a self-scheduler interface. At the top, there is a dark blue header with the "United States Census Bureau" logo. Below the header, there is a paragraph of text: "Please select the day and time combination from below when you are available for a 30 - 45 minute interview with the U.S. Census Bureau. A researcher will send you confirmation of your selected appointment time. All appointments are Eastern (Washington, DC) time zone." Below this text, there is a section titled "Thursday, July 1, 2020" followed by a list of ten radio button options, each representing a different time slot: "Thursday, July 1 at 9:00 am, Eastern", "Thursday, July 1 at 10:00 am, Eastern", "Thursday, July 1 at 11:00 am, Eastern", "Thursday, July 1 at 12:00 noon, Eastern", "Thursday, July 1 at 1:00 pm, Eastern", "Thursday, July 1 at 2:00 pm, Eastern", "Thursday, July 1 at 3:00 pm, Eastern", "Thursday, July 1 at 4:00 pm, Eastern", and "Thursday, July 1 at 5:00 pm, Eastern".

5

Ok, so we had lists of the types of companies – generally and specific – to test the content with. Now we needed a systematic way to recruit companies. On your screen now is a screenshot from a self-scheduler. This is a type of survey that we program where the respondent can select the day and time of their interview, and the number of appointments for any given day and time are limited to the number of interviewers available, such that when a respondent selects an appointment and the quota for that appointment is filled, that specific day and time are no longer available for other respondents to select. The self-scheduler was born out of a need for flexibility at the start of the global pandemic, but we have found that it helps to streamline our recruitment efforts, and it has become part of the standard operating procedure for recruitment for interviews where appropriate.

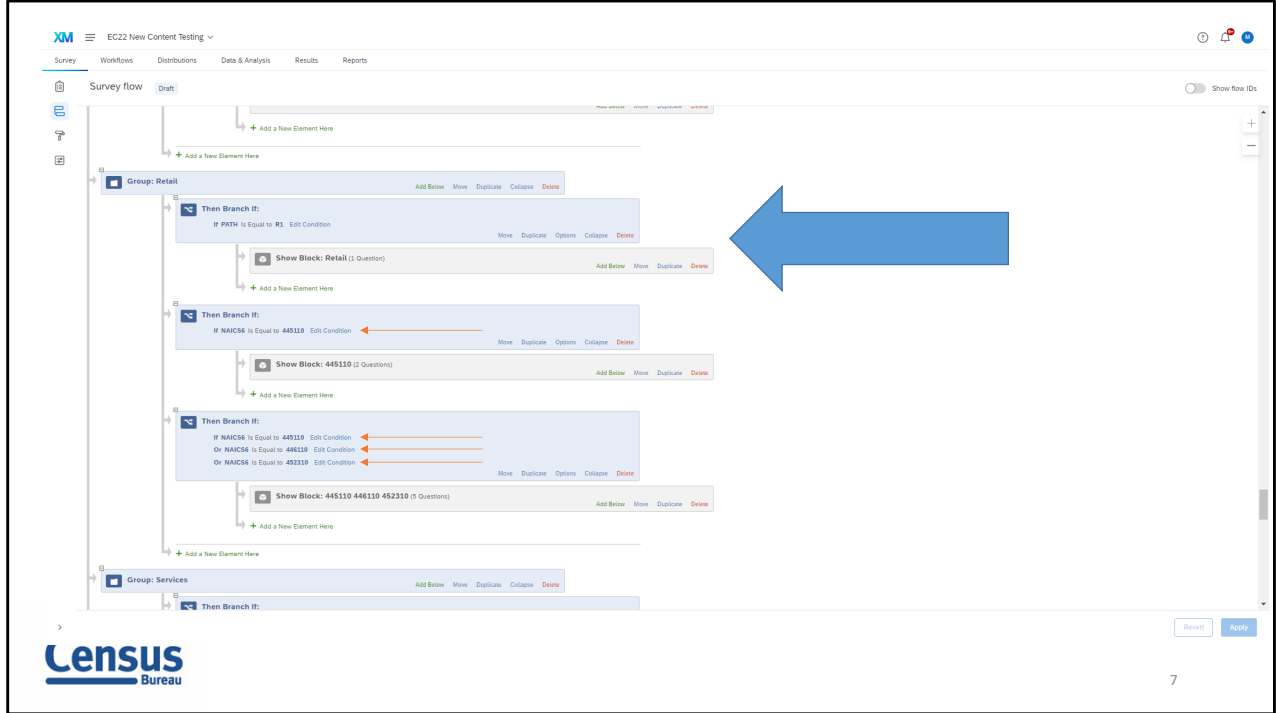
Each week, I tailored the scheduler based on the availability of the interviewers, and then determine which questions needed additional testing in a given week and deploy the scheduler to those targeted companies for the week. This meant that we could plan our interviews about a week out, but no further so that if we were lagging behind with any given set of questions, we could make it up the following week. At the same time, every morning I pulled the interviews from the previous day to provide updated response tracking to help with this weekly targeting of cases. By keeping a close eye on which types of businesses had scheduled interviews and which had completed interviews, we could pivot as the project needed.



Now on to the next challenge – the incredibly complex test instrument for this work. We needed an interactive survey that we could tailor based on various aspects of the company. On your screen now is the Survey Flow – the order and functionality – of the test survey instrument. I know, you can't really read any of it – I am sharing this with you to give you a sense of the size and scope of this test instrument and then we will zoom in and look at a few specific points.

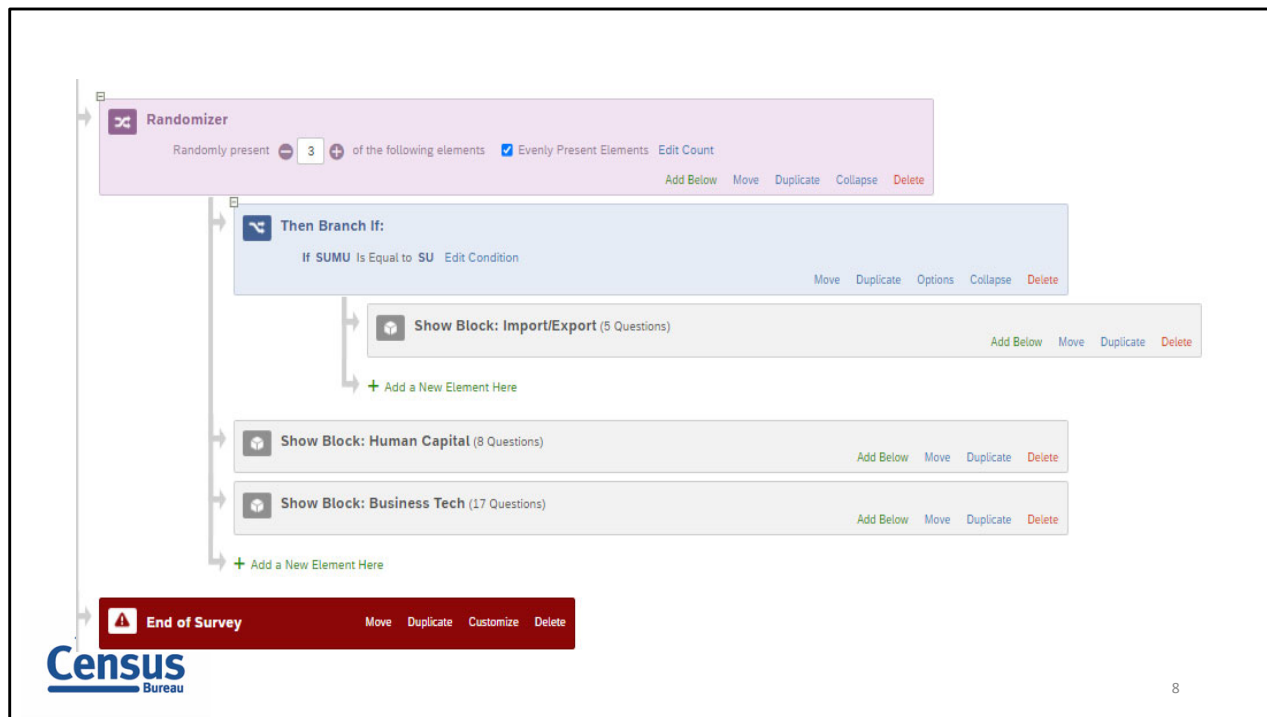
Here is a list of embedded data – these data came from the Business Register and included not only attributes of the company used for routing the instrument, but also contact information, the interview appointment, and other pertinent information needed for testing.

Next up, we needed a way of grouping like questions together. We grouped the test questions largely by the economic sector that would be receiving that question. So, the first group of questions might be for manufacturing, the next retail, the next wholesale, and so on.



Let's zoom in on one of these groups –on screen now is the Retail group. You can see that cases are assigned to specific routes, sometimes as specific as their six digit North American Industry Classification System (NAICS) code.

The first branch here is for a group of retail companies based on several characteristics that we labeled R1 – retail path 1. However as we move our way down the survey flow we get more and more specific, with these questions going only to specific NAICS codes.



Finally, we had some questions that were sector neutral – that is, some of the questions needed to be tested across the economy. To ensure that we had enough cases of each kind, we programmed a randomizer to present blocks of questions on the fly as the respondent worked through the test instrument.

All of this very complicated programming was completely unknown to the respondent – they received a questionnaire tailored to their company, and moved seamlessly through the questions while on the phone with an interviewer.



# Interviewer Protocol

The diagram illustrates the difference between the respondent and interviewer screens for a specific question. On the left, the **Respondent Screen** shows the question: "Item 28 (Self Checkout) Did this establishment offer a way for customers to self checkout, with no or minimal assistance, in 2020?" with radio button options for "Yes" and "No". On the right, the **Interviewer Screen** shows the same question wording but includes a "Probes" section with four bullet points for gathering additional information: "In your own words, what do you think this question is asking?", "How did you come up with your answer? (Would you need to ask others for assistance in answering this question? How is this information kept in your records?)", "What does the term 'self checkout' mean to you? What does 'with no or minimal assistance' mean to you?", and "How easy or difficult would you say this question is to answer? (Time estimate?)". Below the probes is a text box for the interviewer to enter notes. A blue arrow points from the question on the interviewer screen to the question on the respondent screen, indicating the mapping between the two views.

**Respondent Screen**

**Interviewer Screen**

United States<sup>®</sup>  
Census  
Bureau

9

One challenge of the size and scope of this project was the number of interviewers necessary to successfully meet the target goals of the project. To standardize data capture of quotes and feedback, we took the finalized respondent instrument and modified it slightly for the interviewer. On the left side of your screen is a screenshot of a typical question within the respondent instrument. On the right side of your screen is the corresponding screenshot from the interviewer protocol. Notice that the interviewer sees the same question wording as the respondent. However, whereas the respondent sees the corresponding response options, the interviewer has the probes for that specific question embedded on the screen along with a capture box for the interviewer to enter notes on the fly. This way, once the interview is complete, we can pull the notes by question, and can even match up the interviewer notes with what the respondent entered into the test instrument.

## Protocol updates



United States<sup>®</sup>  
**Census**  
Bureau



10

While interviews were being conducted, we also had to plan for extensive in-field support. One of the major ways that we kept this project dynamic and flexible was regular updates to the protocol as we noticed trends in response. On your screen now are two dancing corgis. You may be asking yourself – why are there dancing corgis on my screen? Well, when I would program updates to the protocol, I would embed a different gif at the top of the screen and let the interviewers know which gif they should be seeing for that day. If an interviewer opened a protocol and had, say, singing cats instead of dancing corgis, that was an easy way for them to know that their protocol was out of date.

## In-field support

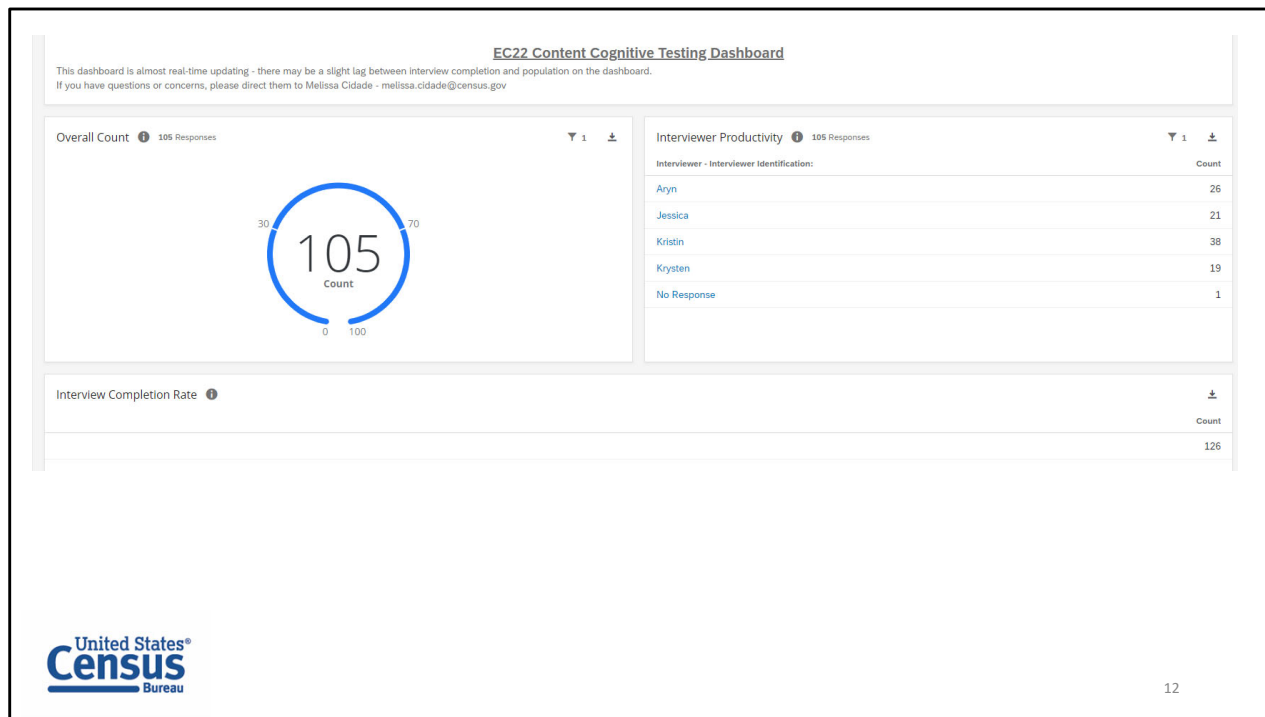
- Project leader – worked with stakeholders, rescheduled interviews, set targets and goals
- Dedicated chat – communication between interviewers and support staff
- Real-time troubleshooting – could answer issues in real- or near-real-time



11

We also had a system for in-field support while interviewers were conducting interviews. First, we designated a project lead – in this case, it was Kristin Stettler. This meant that Kristin interfaced with stakeholders within the Census Bureau to ensure that testing was meeting their needs. She also handled any interview that needed special handling – for example, if a respondent needed to reschedule their appointment, Kristin would work out the logistics for that interview and would often conduct rescheduled interviews herself.

We also had a dedicated chat for interviewers. While we are not authorized to share any protected data over chat, interviewers could use it to alert me to issues with any given case. This chat was helpful for invalid or nonresponsive URLs for both protocol and test instruments, questions from respondents outside of the scope of the current project, and general progress tracking and communications. We found it to be an integral part of the success of this project. And, it allowed for real-time troubleshooting: an interviewer could throw a question in the chat and get an answer from the project lead within moments. This made interviewing smoother and allowed us to deal with issues in real time or near real time.



A major component of this research was keeping a firm grasp on response tracking. We had so many very specific types of businesses we needed to interview that we ended up conducting daily tracking to make sure we were hitting our targets. One way we did this was through a real-time dashboard which provided overall counts of completed interviews. I could also track by interviewer to make sure that no one single interviewer was being overtaxed. Note that the screenshot on your screen now is not from the 'final' dashboard – this was about halfway through the project.

### Critical Item Response Tracking

		N
GENERAL MANUFACTURING, QUESTION 1	Sum	3.00
M1 PATH	Sum	9.00
R1 PATH, QUESTION 1	Sum	11.00
R1 PATH, QUESTION 2	Sum	9.00
R1 PATH, QUESTION 4	Sum	9.00
R1 PATH, QUESTION 5	Sum	2.00
R1 PATH, QUESTION 6	Sum	1.00
S1 PATH, QUESTION 1	Sum	4.00
S2 PATH, QUESTION 1	Sum	5.00
S3 PATH, QUESTION 1	Sum	10.00
S4 PATH, QUESTION 1	Sum	12.00
S4 PATH, QUESTION 2	Sum	NONE
S5 PATH, QUESTION 1	Sum	1.00
S6 PATH, QUESTION 1	Sum	NONE
S7 PATH, QUESTION 1	Sum	22.00
S7 PATH, QUESTION 2	Sum	16.00
GW PATH, QUESTION 1	Sum	10.00
GW PATH, QUESTION 3	Sum	1.00
GW PATH, QUESTION 6	Sum	10.00
W1 PATH, QUESTION 1	Sum	5.00
W1 PATH, QUESTION 3	Sum	NONE
W1 PATH, QUESTION 6	Sum	5.00
W3 PATH, QUESTION 1	Sum	11.00
IA PATH	Sum	7.00

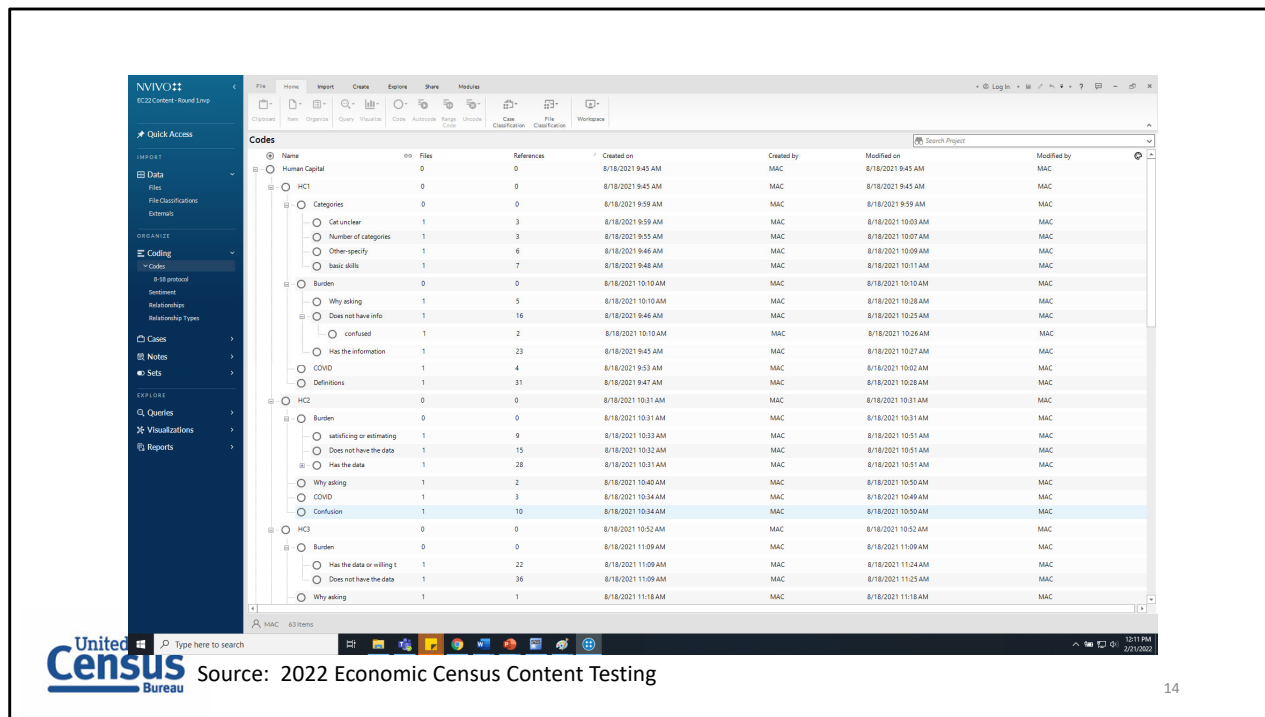


Note: contains fictional data.

Source: 2022 Economic Census Content Testing Documentation

13

Another way I tracked daily responses was through tables. Lots and lots of tables. On your screen now is a main one we looked at each day – these are the critical route points of the instrument, not just did the case meet the requirements to send them down a path, but once on that path, did the responses cover all variations of skip patterns within the path. So, for example, I may have had 11 cases test the Retail 1 (R1) path at question 1, but based on their responses, I had 9 cases then see question 2, and just 2 cases see question 5. In this way, it was not enough to track the number of interviews completed, or even the number of interviews completed by sector....rather, we had to track the number of interviews by key items. Note that these do not represent the final numbers for this project, they are just illustrative.



Finally, we also needed a way to stay flexible with data analysis. This was particularly true for within and between round results – how could we churn through a high volume of interviews and provide quick turnaround impressions of how the questions were testing?

We ended up automating as much as possible the data handling and analysis by importing the notes from the online interviewer protocol directly into a qualitative data handling software program. On your screen now is the coding scheme for a set of questions – this set of codes were developed mid-field, and given their performance, we found the questions to be problematic and ultimately dropped them from subsequent rounds of interviewing. In this way, we could streamline interviews as much as possible and be responsive to findings as they came in.

## Lessons Learned

1. Know as much as possible about your cases
2. Think through the order of operations
3. Lean on technology to keep connected and organized
4. Specialize the tasks –programmer, project lead, and interviewers



15

I have thrown a lot of process at you, and some of this may be able to be replicated based on your project needs, but some of this may be specific to our infrastructure and needs. So, I'd like to leave you with four 'lessons learned' from this project. The first is a bit of an axiom for establishment survey testing anyhow: know as much about your cases as you can prior to beginning an interview. You'll recall at the start of this presentation I showed you the list of embedded data that we were able to pull from the Business Register so that we knew about our cases before getting on a call. This proved invaluable not only to providing context for follow-up questions and probes, but also to identify when an instrument was not performing as expected based on what we knew about the case – for example, a retail case getting questions for a manufacturing company. Knowing about the business before calling is an integral part of managing high volume cognitive testing.

Next, we spent a lot of time upfront, thinking through each step of this project. We carefully laid out a recruitment, programming, interviewing, and analysis plan prior to beginning any work so that we could identify targets and work to meet them. We never had to worry if we were on track or not; we could tell just by looking at the tracking data from that day, or by looking at the dashboard. We had procedures in place for in-field questions, for updates to the protocol, and even for exporting and

analyzing data.

Number three is the biggest takeaway from this project – wherever possible, lean on available technology to keep a large project like this organized and moving forward. We used everything we had at our disposal, from statistical analysis software to an online survey platform, chat functionality, and even qualitative data handling software to keep the project on target but also to maintain flexibility.

Finally, one way that this project is different from most cognitive testing that we do is that we had more specialized roles on this project, and it served the purpose well. Usually, we use a ‘rule of two’ – that is, there are usually two trained interviewers handling all of the recruiting, interviewing, and analysis for a project. For this project, though, we had a dedicated project lead, we had up to five interviewers depending on the round of interviewing, and we had a dedicated programmer and data analyst to launch and then maintain the testing. This specialization – coupled with regular communication both on chat and in standing meetings – meant that each staff member could focus almost exclusively on their contribution to the project while also ensuring that everyone’s impressions and understandings of findings were included.



Thanks!

Melissa Cidade

[Melissa.Cidade@census.gov](mailto:Melissa.Cidade@census.gov)

Kristin Stettler

[Kristin.J.Stettler@census.gov](mailto:Kristin.J.Stettler@census.gov)

Kelsey Drotning

[Kesley.Drotning@census.gov](mailto:Kesley.Drotning@census.gov)

