

Remotely Testing the Integration of Machine Learning in the 2022 Economic Census

Melissa A. Cidade, Ph.D.

Economic Statistical Methods Division (ESMD), U.S. Census Bureau

Wednesday, April 14, 2021

Federal Computer Assisted Survey Information Collection
(FedCASIC) Workshops

Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY21-ESMD002-020)

Shape
your future
START HERE >

United States
Census
2020

My name is Melissa Cidade, and I am a survey methodologist in the Census Bureau's Economic Statistical Methods Division (ESMD). Before I begin, I'd like to recognize Temika Holland and Krysten Messner, as well as the 2022 Economic Census Stakeholders, all from the Census Bureau, for their critical contributions to this presentation.

Data just released from the
2017 Economic Census

The 26,075 Other Services (NAICS 81) establishments in Pennsylvania reported revenue of \$23.0 billion in 2017.

www.census.gov/economiccensus

United States[®]
Census
Bureau

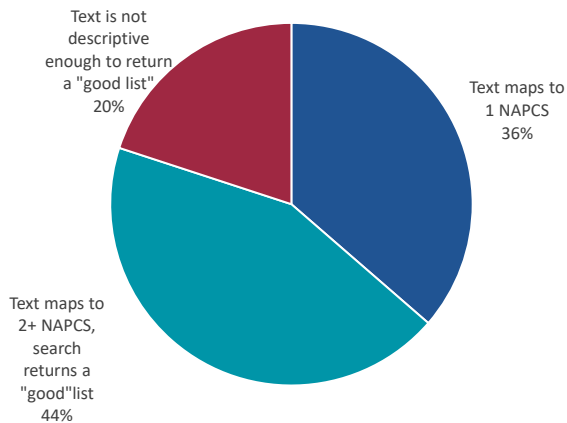
U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov



Like Temika, this research is on behalf of the 2022 Economic Census. Again, this is a data collection effort conducted every five years – a quinquennial data collection – for years ending in 2 and 7. The Economic Census is entirely web reporting, and we only provide paper forms for single unit Island Area companies, so very few.

The 2022 Economic Census – or EC22 for short – will collect data on employment, payroll, and value of goods and services produced in 2022, as well as other topics. The data from EC22 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 program use.

Capturing “other” products



Source: write-in responses from EC17, internal analysis

3 2020CENSUS.GOV

NAPCS Group	Number of Write-in NAPCS	Proportion of All Write-Ins
1 (Mining)	1,700	0.2
2 (Manufacturing)	41,400	4.6
3 (Construction)	34,700	3.8
4 (Wholesale)	99,900	11.0
5 (Retail)	296,000	32.8
6 (Utilities)	3,300	0.4
7 (Services)	138,000	15.3
8 (Tax-Exempt)	48,000	5.3
9 (All Others)	240,000	26.6
Total	903,000	100%

Shape your future
START HERE >

United States
Census
2020

We need to talk about the NAPCS – the North American Product Classification System (NAPCS). NAPCS, like the NAICS codes Temika has talked about, is a nested taxonomy of goods and services. It is not based on industry, but can be linked to industry, and it is consistent across the three North American countries. At its most detailed, NAPCS provides standardized information about the goods and services of companies across the country.

NAPCS is robust and complex system of grouping like products together. Of course, we know no coding scheme is exhaustive. So, we include an other/specify capture box when asking respondents to identify their product codes.

About one MILLION in-scope code-able write-in responses needed to be handled by Census Bureau analysts as a result of the write-ins for EC17. The pie chart on your screen outlines what happened to those almost 1 million written responses in 2017. (click) About 36% were mapped back to 1 NAPCS category - analyst selected 1 NAPCS and made no suggestions of other NAPCS that might be a relevant match to the write-in. (click) Another about 44% are specific enough that they can be mapped to a NAPCS, but not specific enough that only 1 NAPCS is definitely the match. These are cases where multiple NAPCS could match a relatively specific term (like "lumber" or

"ambulance") or it could be a multi-segment, where the respondent has reported all of their products/services in one write-in. (click) Finally, about 20% were either too broad to return a response (for example, sales, general merchandise, etc.) or the write-in was a remark/comment (i.e. the respondent is stating that they have been put in the wrong industry, making a comment on the survey as a whole, etc)

(click) This table shows you the count data by Sector. Almost a third of in-scope write in responses came from the retail sector alone. That is time and resource intensive, and we believe that better survey design could lead to a decrease in write-in responses. We worked with other researchers, stakeholders, economists, programmers, and others within the Census Bureau to design two versions of the form to take out for usability testing with the intention of decreasing the write-in responses for this question. These new versions of the form included a new machine learning algorithm to help respondents select NAPCS classification codes on the fly as they identified products and services that were not pre-listed on their forms.



However....then, 2020 happened. Our plans for visiting businesses to see how they interacted with the versions of the forms were put on hold as we went into mandatory telework and federal employee travel prohibitions.

We did, however, seen an opportunity to test out new methods of remote usability testing.

Remote Testing Research Protocol



5 2020CENSUS.GOV

Shape
your future
START HERE >

United States
Census
2020

After some trial and error on this and other projects, we settled on the following research protocol to conduct usability testing remotely.

(click) We start with a recruitment email to within-scope firms. This email requested respondents to schedule an interview with a Census Bureau researcher.

(click) For those that did schedule an interview, we then send them a follow-up email confirming their selected appointment, describing that the interview will include using Skype for Business, and affirming their interest in participation.

(click) The morning of the interview, we sent out an email that again confirmed the appointment time and date, and providing a link to complete an online consent form.

(click) Thirty minutes before the interview, we sent a final email containing the interview call-in information, the link to the test instrument we would be using with the respondent, and providing the consent form URL again if the respondent had not already submitted one online.

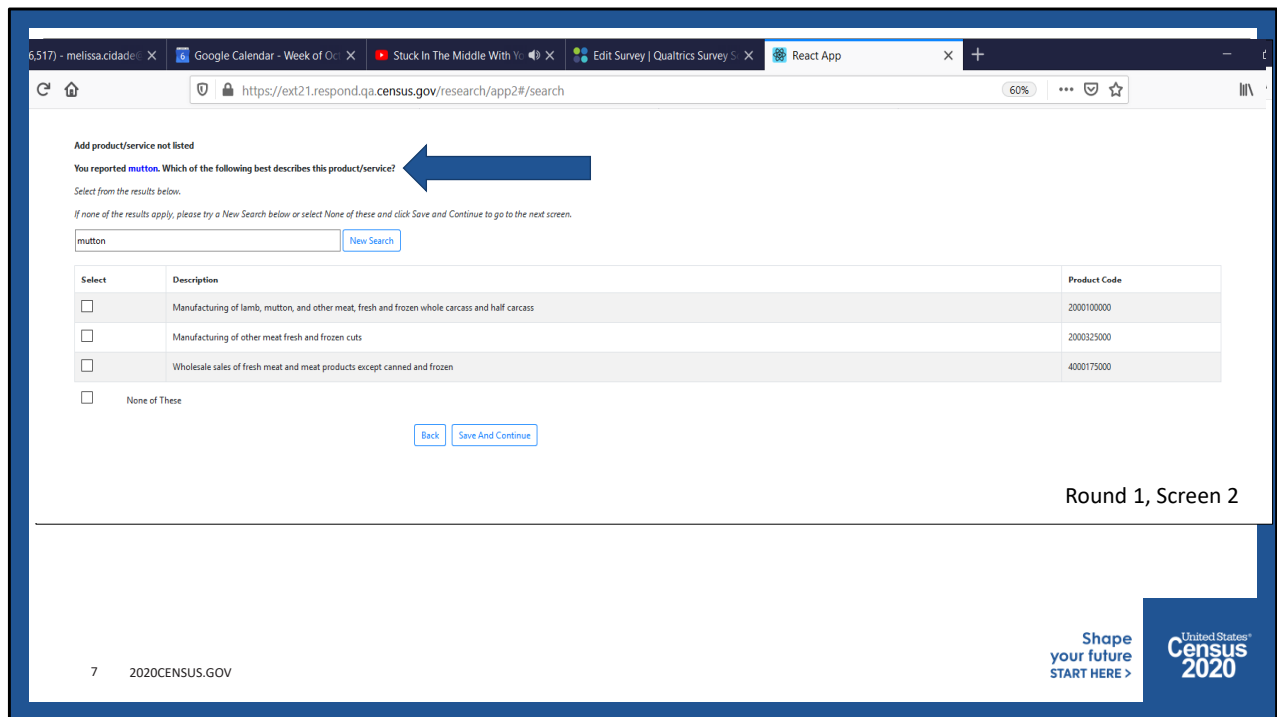
(click) From there, we conducted our interviews using Skype for Business so that we could have respondents share their screens with us to see how they were interacting with the test instrument.

Respondent Overview

	Round 1	Round 2	Total
Service	9	20	29
Wholesale/Retail*	25	3	28
Manufacturing	11	7	18
Mining/Construction*	5	6	11
Total	50	36	86

*Industries collapsed to meet disclosure requirements.

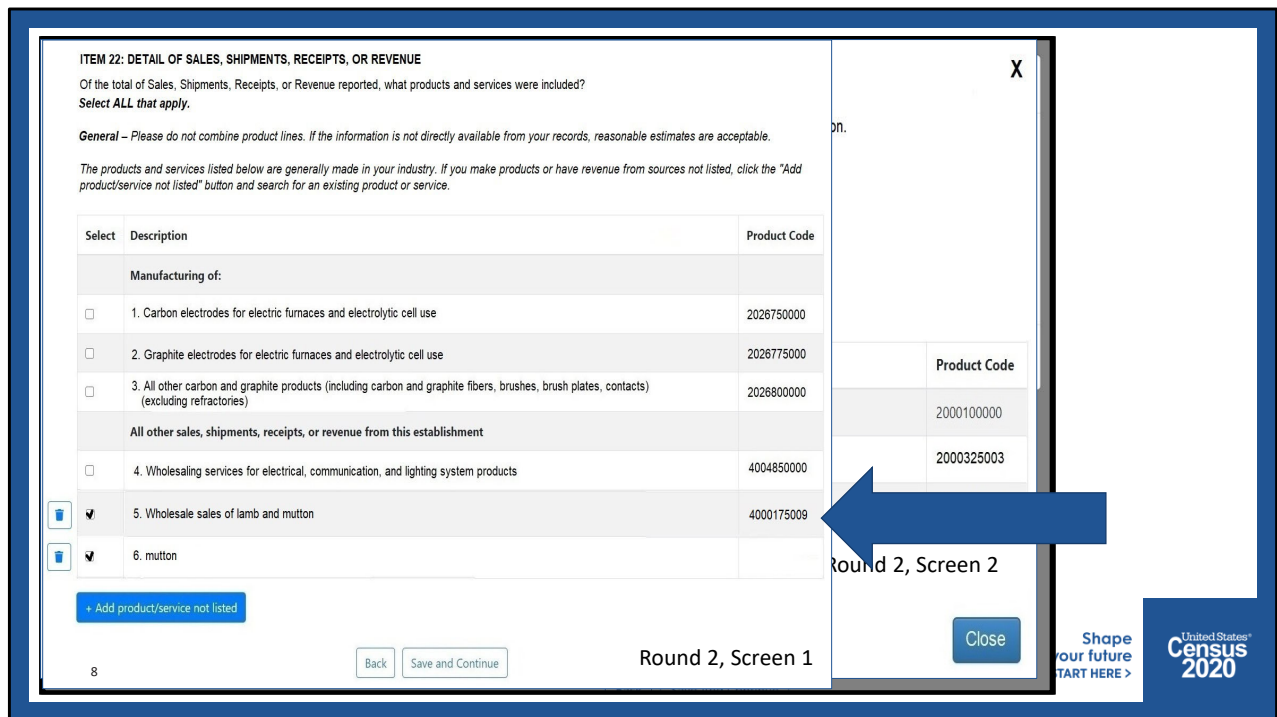
In total, we conducted 86 interviews between August and November 2020 over two rounds of interviewing. We recruited the first round of respondents from a random list of multi-unit firms sampled for the 2017 Economic Census. In the second round, we recruited from firms identified as having ‘written in’ a response to the NAPCS question on EC17. Interviews lasted no more than 30 minutes, respondents and interviewers used Skype for Business to complete the interview. We also invited stakeholders from the EC22 team – including programmers, statisticians, economists, and project managers – to observe the interviews using Skype for Business.



For the round 1 testing, we tested two versions of screen 1, the “product listing screen.” Screen 1 version 1, on your screen now, features an “add products or services not listed” button.

(click) and here is screen 1, version 2, (click) which features three open ended “other, please specify” captures. This was closer to the actual form from 2017; respondents would enter their text into the capture box, and upon clicking “save and continue,” they would be directed to screen 2, the “search results screen.”

(click) which you see on your screen now. (click) At the top, you’ll notice that the respondent write-in is prefilled, and (click) the ‘best matches’ from the machine learning algorithm tries to match up with the appropriate NAPCS code.



For round 2, we dropped the second version of the first screen – the one with the open boxes. We also retained the bottom buttons for navigation, and instead, when the respondent clicked on “Add product/service not listed”, the search was automatically execute.

(click) We also started to field test more robust instructions at the top, especially drawing attention to the search functionality.

(click) We gave screen 2 – the results screen – a bit of a makeover, too. Notice that once the respondent clicks to add, a modal window (not a popup!) appears as an overlay to the first screen. There, the respondent can type in the additional product or service missing from his prelist.

(click) and then, once the respondent clicks search, he is presented with a list of products that the machine learning algorithm has identified as most likely to match.

(click) Once the respondent enters all of his additional products, he is redirected back to screen 1 with his additional products at the bottom of his list, and he can continue through the rest of the instrument.

Example of Intended Interaction

Industry: Service

Now, let's watch a quick video of an "ideal type" response for round 2. This video does not have sound, so I'll provide some narration.

Key Takeaways

1. When integrating machine learning into your survey, test multiple versions of screens, make search functionality clear and automatic, and include instructions.
2. Keep navigation buttons for navigating through the survey; include additional buttons for additional functionality.
3. 2020 was challenging but also afforded some opportunities for methodological innovation.

Ok, and in summary, here are the three things you can take away from this presentation.

The first is that when you are testing a complex instrument that is integrating machine learning functionality, be sure to have multiple versions of layouts and functionalities to take into the field with you to test. You may want to make the search functionality automatic – but be sure to clearly communicate to respondents that clicking a button will execute a search. And, instructions cue respondents to the task at hand on screen.

The second is to design the buttons purposefully – keep navigation buttons like “back”, “forward,” or “submit” strictly for navigating, and incorporate additional buttons for each additional functionality you develop.

Finally, 2020 was a hard year for survey methods (and everyone), but it was also a chance to try out new technologies and methodologies. My team was flexible and innovative, and in the end, we were able to complete usability testing completely remotely, including the addition of the self-scheduler, the use of screen capture, and others.

Thanks!

Melissa A. Cidade, Ph.D., Survey Methodologist
Data Collection Methodology & Research Branch
Economic Statistical Methods Division
U.S. Census Bureau
Phone: 301-763-8325
Email: Melissa.Cidade@census.gov