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

2020 CENSUS PROGRAM INTERNAL MEMORANDUM SERIES: 2019.14.i

Date: April 23, 2019

MEMORANDUM FOR: The Record

From: Deborah M. Stempowski **(signed April 23, 2019)**
Chief, Decennial Census Management Division

Subject: 2020 Census Evaluation: Analysis of Census Internet Self-Response Paradata by Language Study Plan

Contact: Jennifer Reichert
Decennial Census Management Division
301-763-4298
jennifer.w.reichert@census.gov

This memorandum releases the final version of the 2020 Census Evaluation: Analysis of Census Internet Self-Response Paradata by Language Study Plan, which is part of the 2020 Census Program for Evaluations and Experiments (CPEX). For specific content related questions, you may also contact the authors:

Renee Ellis
Center for Behavioral Science Methods
301-763-5122
renee.ellis@census.gov

Joanna Fane Lineback
Center for Behavioral Science Methods
301-763-7599
joanna.fane.lineback@census.gov

Rodney Terry
Center for Behavioral Science Methods
301-763-5475
rodney.terry@census.gov

Patricia Goerman
Center for Behavioral Science Methods
301-763-1819
patricia.l.goerman@census.gov

Anna Sandoval Giron
Center for Behavioral Science Methods
301-763-3575
anna.b.sandoval.giron@census.gov

United States Census 2020

2020 Census Evaluation

Analysis of Census Internet Self-Response Paradata by Language

Study Plan

Renee Ellis, CBSM
Fane Lineback, CBSM
Rodney Terry, CBSM
Patricia Goerman, CBSM
Anna Sandoval Girón, CBSM

United States™ Census Bureau

4/23/2019
Version 3.1

Page intentionally left blank.

Table of Contents

Page intentionally left blank.

I. Introduction

For the first time, the decennial census will extensively offer internet reporting. The internet instrument will be available in 12 languages, including English, Spanish, Chinese, Vietnamese, Korean, Russian, Arabic, and Tagalog. Although the 2010 Census included an online version, it was only available in English, and because it was unadvertised and limited to only the short form, only 63,000 households responded online (Whitworth 2002). In 2020, there will be full advertisement of the online form and community partnerships will encourage use of the online form.

This new online instrument will provide a unique opportunity to study non-English language reporting. Previously, we had little, if any, prior information about internet reporting for many of the 12 languages besides English and to a more limited extent Spanish. Until now, we certainly have nothing on the scale of the decennial census that we have been able to study. Given the scale—in terms of both population and languages covered—of the 2020 Census, what we learn can be used not only to evaluate the 2020 Census and make improvements for the 2030 Census, but also to inform the numerous surveys and censuses that offer or plan to offer reporting in multiple languages.

Internet paradata includes information about computer settings and user actions within the instrument, including the timing of those actions. Commonly collected items include browser type, device type, operating system, login, logout, help access, answer change, warning message, error message, and page information.

Web paradata can be used for many purposes, such as to identify usability issues and problematic questions. They may be used to evaluate the ease-of-use of the instrument during high-traffic times. For non-English languages, the paradata may offer insight into the ease of use across languages, any translation issues, and the use of built-in translator tools (such as Google translate) versus the census provided translations. They may be used as a confirmatory tool if they are combined with the findings from qualitative language studies. They web metrics may be compared with paper metrics to evaluate the effects of things such as the Census Questionairre Assistance (CQA) across languages.

Up-to-now, language work had been largely qualitative in nature (see section III for more detailed information about qualitative language work). We are not aware of any previous language research using internet paradata. Although previous research has resulted in recommendations for non-English content and translation, as well as procedures to increase cultural and linguistic appropriateness, ongoing research is necessary to understand how respondents actually use instruments in various languages and how that affects data quality. This is especially timely now, because of increases in usage of online surveys.

We will examine 2020 Census web paradata by language. The paradata can be used to identify usability or problematic questions specific to certain languages. The paradata findings can be compared to previous qualitative findings to look for any gaps.

II. Assumptions

- There will be space to store and analyze data.
- Paradata needed for research will be collected and made available, including answer changes, instrument language and toggling, and browser language setting.
- Paradata will have a flag to distinguish CQA from self-response.
- Paradata or response data will have a flag to distinguish between respondents who completed the online form with an ID and those that completed non-ID.
- 2020 Census data will be made available.
- Census test web paradata will be made available.
- Previous 2020 qualitative pretesting and usability languages study reports will be made available.
- Results and reports for external audiences will go through the required review process, including review by Data Products and Dissemination.

III. Background

Previous cognitive research in languages other than English

In support of obtaining high quality census data from hard-to-count, limited-English-speakers, the Center for Behavioral and Survey Measurement (CBSM) has conducted non-English questionnaire design and pretesting research for over 15 years. This research has included expert reviews, cognitive interviews and usability testing of automated instruments, focus groups in non-English languages, and doorstep interaction observations. There is little, if any, existing research using paradata to study non-English responses.

As part of 2010 Census evaluations and experiments, CBSM researchers observed doorstep interactions in seven languages (Spanish, Chinese, Vietnamese, Korean, Russian, Arabic and Portuguese) to obtain basic information about how these interviews were conducted. For another 2010 evaluation, CBSM researchers overserved nonresponse follow-up (NRFU) of English and non-English speaking Hispanic households in Texas. Researchers observed that, of 23 Hispanic households interviewed by NRFU enumerators, 10 interviews were conducted in English, eight in Spanish, two in English and Spanish, and two with the enumerator speaking English and a household member serving as an interpreter.

As part of recent census tests, CBSM researchers observed non-English doorstep interactions and messaging and reviewed field enumerator non-English materials. In addition, CBSM researchers have studied the topic of low literacy respondents and how they interact with surveys and forms

Pretesting of non-English census questionnaires over the years has uncovered a number of comprehension and translation issues. Cognitive interviews before the 2010 Census found differences in the difficulty of filling out paper census forms by language. Some of that work resulted in changes to the instruments. Additional multilanguage pretesting was

conducted in the decade leading up to the 2020 Census. That work was also used to improve translations. Tests preceding 2010 found that English speakers generally had few problems with navigation and tended to skip instructions (Pan et al, 2009). English speaking participants generally found the questionnaire language to be standard and had few problems. The few problems that were identified were generally in the explanations of who is considered a household member, rental/ownership status and less common relationships like roomer/boarder or housemate. Participants who spoke languages other than English tended to have difficulty in the same places as English speaking participants, they also had additional places where things were difficult.

In research prior to 2010, Russian research showed similar findings to English. Participants felt the form was routine and basically felt the language was standard (Pan et al, 2009). Participants who spoke Russian had the most trouble with the rent/own question. There were some additional problems with understanding some particular terms like foster and adopted child and some participants believed the survey was asking about all people in the apartment building rather than just the home.

Participants who spoke Chinese, Korean, or Vietnamese had more trouble with the form (Pan et al, 2009). This largely stemmed from translation and cultural issues. Forms in these languages tended to use more formal translations rather than modern words and the words were often ordered in the same way as the English version rather than in a way that flowed for the language. There were also issues with terminology that meant something different across languages or did not have an equivalent, terms like mobile home or foster child. The questions that were most difficult were those about who to include in the household and rent/ownership of the home. Age and name questions were also more difficult because of differences in counting and naming conventions. Some additional relationship categories like unmarried partner were also more difficult. Chinese and Vietnamese participants tended to have difficulty navigating the form and understanding the purpose of the form (Pan et al., 2009). For Korean respondents, difficulty navigating was dependent upon age and education, older, less educated participants had more difficulty (Pan et al 2009, Park, Pan & Sha, 2009).

There has been additional cognitive testing leading up to the 2020 Census. Contractor testing reports have not all been posted to the Census Burea website but they are available internally. They seem to indicate a few continuing issues with translations and navigation. Some of these translations have been updated since the tests and this will allow us to evaluate the differences in the online paradata between the tests and the 2020 Census. Expert reviews and cognitive interviews were conducted in Chinese, Vietnamese, Korean, Arabic, and Russian (Sha et al. 2016-1). Expert reviews, cognitive testing and usability testing was conducted in Spanish (Sha et al. 2016-2). The expert reviews found that the translations for 2010 were mostly effective and they recommended only a few changes. The changes recommended by the expert reviews were then cognitively tested. Findings from the first round of testing were then tested in a second round.

For all languages where testing was done of online instruments, respondents had difficulty finding and using their online identification number (Sha et al. 2016-1, Sha et al. 2016-2). Research has also found that Spanish-speaking respondents in the U.S. may have limited ability to use mobile devices, even if they have access (Garcia Trejo & Schoua-Glusberg, 2017). Also Spanish-speaking respondents in the U.S. have difficulty with procedures to log on and navigate web-based surveys (Lykke & Garcia Trejo 2018). These issues may not be limited to Spanish speakers. Respondents who participated in cognitive testing in Arabic, Korean, and Russian were more familiar with online surveys and had a relatively easy time navigating the instrument. Respondents who answered in Chinese, Spanish, or Vietnamese were less familiar with online navigation and had more difficulty (Sha et al. 2016-1).

The most common area for difficulty in all languages was in understanding residency rules and instructions for questions that included definitions. Additionally, respondents answering in Chinese and Spanish had difficulty with the undercount questions and the race/ethnicity questions (Sha et al. 2016-1, Sha et al. 2016-2). Korean, Vietnamese, and Russian respondents had difficulty with the county/township question (Sha et al. 2016-1). Korean respondents tended to overcount college students. Respondents answering in Arabic had trouble with the new MENA category (Sha et al. 2016-1).

IV. Research Questions

We ask the following questions:

- Can we identify usability issues or problematic questions specific to certain languages?
- Does the paradata confirm qualitative language research findings?
- Does the use of a CQA agent mitigate difficulty for respondents who use a language other than English? How do CQA responses differ from self-responses?

We do not currently have access to the non-English instrument language versions of the online instrument, nor have we done detailed languages paradata analysis for other surveys, so we are not quite sure what differences to expect, although we certainly expect there will be differences. Based on limited qualitative research, we hypothesize respondents who use the non-English instruments may have increased difficulty with some questions. We hypothesize that the citizenship question and race ethnicity questions will produce higher nonresponse rates for respondents who use non-English forms because of potential sensitivity. We hypothesize that some non-English language respondents will have more difficulty logging on and navigating the instrument than English respondents. Cognitive interviews suggest this is particularly true of Spanish and Chinese speakers. We hypothesize respondents who use non–English forms may be more likely to come in as non-ID's because of difficulty identifying their ID number compared to English. Finally, we hypothesize that questions identified in cognitive research as more problematic will be more difficult for respondents, these include undercount, age, name, and rent/own.

V. Methodology

A. Evaluation design

We will use the 2020 internet paradata, 2018 test internet paradata, 2017 internet test paradata, 2015 internet test paradata, non-English language and English language qualitative reports, and 2020 Census response data to conduct this research. For 2020 ISR and the two most recent census tests, we will start with the descriptive statistics. (See https://collab.ecm.census.gov/div/csm/intranet/Pages/Other-Resources.aspx.). We will calculate the statistics overall and by reporting language. For write-in responses, we can calculate basic statistics, but we would enlist the help of translators to further explore any apparently problematic questions. Below are some examples of the types of statistics we will examine (some are census-specific and not discussed in the previous document):

- What is the percentage of successful logins by language? Percentage of failed logins by language? Percentage of breakoffs by languages?
- How often are users toggling between languages using the census provided toggle? What pages are most frequently accessed in a language other than English? What are the differences between languages?
- How do indicators such a time-in-instrument, time-on each page, time on each question, number of answer changes (in instrument, by page by question), number of times *help* was accessed (in instrument, by page by question), and number of warning messages (in instrument, by page by question) vary by language instrument? We intend on doing basic analysis and comparisons on all questions and pages. However, we will pay particular attention to questions that cognitive interviews indicate are more problematic for someone using a particular language.
- How does the frequency of answer changes vary by language? When answers where changed in open ended questions, what were they changed from? What were they changed to? (data may not be available for last two questions) If we are unable to see changes to answers we will still look at the final write in response for respondents who changed answers.
- How do respondents using different languages vary in type of device and browser used? Mobile versus Personal Computer (PC)? Type of web browser? What is the distribution of browser language settings? Do indicators (logins, errors time in instrument etc.) vary by browser or mobile/PC?
- How does time-in-instrument vary between high and low traffic times?
- Does non-ID and ID reporting differ between languages?

Until 2020 data become available, we will begin our analysis with the 2017 and 2018 census test data and earlier qualitative reports. We plan to look at every question, especially since some preliminary, unpublished research on paradata finds that trouble areas in online instruments do not always match the trouble areas in cognitive testing. We will start by calculating descriptive statistics using the census test data. Although the test paradata do not do not include all 12 languages, we can use the information to begin evaluating changes made based on qualitative findings. Once the 2020 are available, we will add the additional languages to our analysis.

B. Interventions with the 2020 Census

This project has minimal impact on the 2020 Census. The only requested change is that the systems be modified ahead of the census to collect the answer changes as part of the internet paradata. Otherwise, we should have the data we need.

Below is a list of the high level activities for this study:

1. Obtain data paradata from previous 2020 census tests.
2. Obtain English and non-English language qualitative testing reports (e.g., cognitive and usability).
3. Develop initial programs using 2018 ISR paradata.
4. Obtain 2020 Census data and paradata.
5. Analyze data, including comparison to previous qualitative findings.
6. Write report.

C. Implications for 2030 Census design decisions and future research and testing

The proposed research and any recommendations that may come from it are based on the assumption that the 2030 Census will transition to an all-electronic census. Paradata analysis across multiple instruments will help us identify non-English language issues with all survey questions.

VI. Data Requirements

We will use census test paradata from all available tests and decennial paradata when it comes available.

Below is a list of data requested for this research-

Data File/Report	Source	Purpose	Expected Delivery Date
2015 ISR paradata, metdata	National Content Test 2015	analysis	ASAP
2015 ISR paradata, metadata	Savannah Test 2015	analysis	ASAP
2016 ISR paradata, metadata	2016 Census Test	analysis	ASAP
2017 ISR paradata, metadata	2017 Census Test	analysis	ASAP
2018 ISR paradata, metadata	2018 Census Test	analysis	ASAP
2020 ISR paradata, metadata	Deccenial Census	analysis	After data collection as soon as data is available
2020 ISR response data, metadata	Deccenial Census	analysis	After data collection as soon as data is available

Detailed 2020 Census ISR paradata requirements:

- Case-level identifier.
- Time stamp corresponding to each event (each action the respondent made).
- Type of event that occurred (login, entry, exit, submit, field_change, next_action, previous_action, hyperlink, etc.).
- Page in the instrument on which an event occurred.
- Name of the field within a page on which an event occurred (filled when type = field_change).
- Answer that was selected or entered (filled when type = field_change).
- Person-level identifier for household surveys (not sure, if this is populated for other surveys).
- Web address of the link that was clicked (filled when type = hyperlink).
- Vertical screen resolution in pixels (filled when type = login).
- Horizontal screen resolution in pixels (filled when type = login).
- User agent string that provides information on the device, operating system, browser, etc. (filled when type = login).

What was not collected for the Census-eCase test, but needs to be collected in 2020:

- **All information written into any field that is not a check box** (e.g., including birthdate, address, etc.).
- **Every answer change**, including the value it was changed from and the value it was changed to-both write-ins and check boxes.
- **Language instrument being used at any given time.**
- **Browser language setting**, which should be collected with other server information such as screen size and user agent string.
- **Information on language help**. There are some languages that have an instrument and others that do not have an instrument but do have a language specific help file, so it is important to know help file language being accessed.

VII. Risks

Not collecting all of the internet paradata needed for research. Relatively speaking, paradata are a much cheaper alternative than meeting with respondents directly. We cannot stress enough how important it is to collect all of this information. Internet paradata have the potential to be a treasure trove of information, but we must collect them to be able to use them. We have heard that there are no plans to collect answer changes, including for write-in boxes. If this data is not collected then this would eliminate our ability to answer some of our research questions. We continue to work with Decennial to push for collection of this data.

VIII. Limitations

Because this is the first time we have collected non-English decennial census data in an online format, we are limited in how much we can say about the source of different issues

we find by language. In the above section we have included some qualitative findings that we have used to form our hypotheses, but there have not been large scale studies using the paradata to examine language.

IX. Issues That Need to be Resolved

We need to make sure we are planning on collecting the paradata needed to conduct these evaluations. We have been told that there are not plans to collect all possible web paradata. This would affect this evaluation, as well as other evaluations. Additionally, we need access to the data outlined above.

X. Division Responsibilities

Division or Office	Responsibilities
CBSM	• Manage project • Analyze paradata • Write up findings
DSSD	• Provide data • Consult on census response data and paradata • Provide project funds • Provide non-English language translation resources as needed

XI. Milestone Schedule

Evaluation and research or testing study Milestones	Date
Obtain data from all sources except 2020 Census response and ISR paradata	5/31/2019
Merge data; Create test files	7/31/2019
Conduct preliminary analyses using census test paradata	12/31/2019
Obtain access to 2020 Census data	TBD
Obtain access to 2020 Census ISR paradata	TBD
Consult with language experts and translators as needed	ongoing
Language ISR paradata analysis complete (assuming data received at least a year ahead of time)	3/31/2021
Give results briefing (assumes data are available in time to meet earlier deadlines)	5/13/2021
Draft Final Report	6/17/2021
Distribute Initial Draft Paradata by language Report to the Decennial Research Objectives and Methods (DROM) Working Group for Pre-Briefing Review	mm/dd/yyyy
Decennial Census Communications Office (DCCO) Staff Formally Release the FINAL Paradata by language Report in the 2020 Memorandum Series	mm/dd/yyyy

XII. Review/Approval Table

Role	Approval Date
Primary Author's Division Chief (or designee)	mm/dd/yyyy
Decennial Census Management Division (DCMD) ADC for Nonresponse, Evaluations, and Experiments	mm/dd/yyyy
Decennial Research Objectives and Methods (DROM) Working Group	mm/dd/yyyy
Decennial Census Communications Office (DCCO)	mm/dd/yyyy

XIII. Document Revision and Version Control History

Version/Editor	Date	Revision Description
0.1 / RE	8/20/2018	Initial full first draft
1.0	12/6/2018	Draft for DROM

1.1		DROM Feedback
1.2	1/17/19	Project lead feedback and changes
2.0	2/6/2019	Revisions for DROM
3.0	3/14/2019	Revised after process review comments and Drom meeting
3.1	4/23/2019	Copy edited

XIV. Glossary of Acronyms

Acronym	Definition
ADC	Assistant Division Chief
DCCO	Decennial Census Communications Office
DROM	Decennial Research Objectives and Methods Working Group
DSSD	Decennial Statistical Studies Division
EXC	Evaluations & Experiments Coordination Branch
IPT	Integrated Project Team
R&M	Research & Methodology Directorate
CBSM	Center for Behavioral Research Methodology

XV. References

Ajrouch, Kristine, Yuling Pan and Lubkemann. 2012. "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Arabic Repor.t" *Survey Methodology 2012, 05*.

Decennial Statistics Studies Division. 2016 @020 Census Research and Testing 2012 National Census Test Contact Strategy, Optimizing Self-response.

Garcia Trejo, Y.A. and Schoua-Glousberg, A. (2017). "Device and internet use among Spanish dominant Hispanics: Implications for web survey design and testing." Survey Practice 10:3.

Hogan, H., Cantwell, P., Devine, J., Mule, V.T. and Velkoff, V. (2013) Quality and the 2010 Census. Population Research and Policy Review, Vol 32, No. page 637-662.

Isabelli, Christina, Yuling Pan and Lubkemann. 2012. "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Spanish Report." *Survey Methodology 2012, 06*.

Lykke, L., Garcia Trejo, Y. A. (2018)" Results from the Usability Testing of the Spanish Language Version of the 2020 Census Barriers, Attitudes and Motivators Survey (CBAMS)" Survey Methodology #2018-19. https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/rsm2018-19.pdf.

Matthews, K., Phelan, J., Jones, N., Konya, S., Marks, R., Pratt, B., Coombs, J., Bentley, M. (@017). Overview of 2015 National Content Test Race and Ethnicity Analysis Report. https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf.

Elizabeth Nichols, Erica Olmsted-Hawala, Mikelyn Meyers, "Evaluating the Decennial Census Reinterview Option: Results from an Interviewer Debriefing of the 2018 Census Test," CBSM Research Report Series, Forthcoming.

Pan, Yuling, Stephen Lubkemann, "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Evaluation Report," 2013, https://www.census.gov/srd/papers/pdf/ssm2013-02.pdf, **accessed on 7/16/18.**

Pan, Yuling, M. Mandy Sha, Hyunjoo Park, Alisu Schoua-Glusberg. (2009). "2010 Census Language Program: Pretesting of Census 2010 Questionnaire in Five Languages" Survey Methodology #2009-01).

Park, Hyunjoo, M. Mandy Sha, Yuling Pan. (2009) "Cognitive testing as a method of pre-testing questionnaires in high and low context cultures: a comparison of Korean and English cognitive interview data" AAPOR.

Rodrigues, Isabel, Yuling Pan and Lubkemann. 2013. "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Portuguese Report." *Survey Methodology 2013, 15*.

Sha, M., Park, H. , Pan, Y.,Son, J., Schoua-Glusberg, A., Haps E., Kenward K. (2016) "Multilingual Research for 2020 Census Questionnaire: project 4 report" Internal unpublished report.

Sha, M., Schoua-Glusberg, A., Gerber, E., Kenward, K., Engle, J., Glusberg, D., Corado, D. (2016) "Census 2020 Multilingual Research: Project 2 English and Spanish Cognitive and Usability Testing" Internal unpublished report.

Sha, M., Son, J., Pan, Y., Park, H., Schoua-Glusberg, A., Tasfaye, C., Sandoval Giron, A., García Trejo, A., Terry, R., Goerman, P., Meyers, M., and L. Lykke. 2018. "Multilingual Research for Interviewer Doorstep Messages Final Report." Research Report Series. 2018.

Shepherd, Robert, Yuling Pan and Lubkemann. 2012. "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Chinese Report." *Survey Methodology 2012, 08*.

Whitworth, Erin, (2002) "Internet Data Collection." Census 2000 Evaluation A.2.B https://www.census.gov/pred/www/rpts/A.2.b.pdf.

Yoon, Kyung-Eun, Yuling Pan and Lubkemann. 2012. “Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Korean Report.”. *Survey Methodology 2012, 07*.