

### **X's Version**

We recommend a combination of usability evaluation techniques for the Census 2020 instruments:

- Westat's user experience specialists will review the instruments.
- Focus groups will explore users' reactions to the instruments.
- For each instrument, usability laboratory tests will reveal usability issues.

**Expert review.** Westat's user experience team will review English translations of the Census 2020 instruments, even before we recruit any respondents in this project. The reviews will identify potential usability issues. The reviewers need not write a formal heuristic review report.

The findings of those reviews will be helpful when we write the protocols that we will need for our evaluations, including the protocol for the user interviews, the protocol for the focus groups, and the vignettes for the usability laboratory testing. We will ensure that these protocols take into account all of the potential usability issues that our expert reviewers identified.

**Focus groups.** The Statement of Work calls for focus groups, which we will conduct in Puerto Rico. The focus groups will help us to evaluate the Census 2020 instruments. For example, for our evaluation of the Internet self-response instruments, the focus groups will help us learn whether the respondents understood how to enter data about themselves and others in their households accurately.

During the focus groups, the moderators will present sections of the Census 2020 instruments on a large screen, so that all participants can see them. For example, the moderators could present data entry screens, or particular questions or instructions from the instruments. We will be certain to include the sections of the instruments in which our expert reviewers identified potential usability issues.

The focus group participants could then discuss these sections of the instruments. The moderator will ask about the potential ways the questions, instructions, or navigation paths could create confusion. The participants will discuss their approaches to using the instruments, and their reactions to the design of the instruments. The moderator, or the focus group participants, may propose alternate designs. The focus group could then discuss those alternatives. The insights from these focus group discussions could point to effective ways to redesign parts of the Census 2020 instruments.

**Usability laboratory tests.** We will perform user experience evaluations in Westat's usability laboratory. We could also conduct these evaluations in conjunction with the cognitive interviews away from Westat's usability laboratory. Of course, when we conduct usability laboratory tests away from Westat, we may not have the video recording capabilities and the one-way mirror that we enjoy in our own usability laboratory. We could nonetheless conduct adequate tests, for example, at any focus group facility.

We would ask the respondent to carry out a series of vignettes using the Census 2020 instruments. For example, when we test the Internet self-response instrument, we will give the respondent a list of vignettes. The respondent will carry out the vignettes one at a time. A vignette might be "Please pretend you are a man who lives alone in an apartment. You travel to Korea often. You live in Korea about 6 months of every year. Please complete the Internet self-response instrument." Another vignette might be, "Please pretend you and your husband share a

house. Your daughter is away at college most of the year. She lives in a dormitory while she is at college, but she lives with you and your husband every summer. Please complete the Internet self-response instrument.” Another vignette might be, “Please pretend you and your wife live in one house in one city for half the year, and a different house in a different city for the rest of the year. Please complete the Internet self-response instrument.”

As the respondent carries out the vignettes, we could record the screen of the mobile device, using our apparatus for that purpose. We could also record the respondent, using the cameras in the usability laboratory. The recording would have a picture-in-picture image, with the large picture being the screen of the mobile device, and the small picture, embedded in a corner of the image being the user. Alternatively, we might decide that it will be sufficient for purposes of this testing to record only the user, without the screen of the mobile device.

For each vignette, we would observe the respondents’ accuracy. For example, for the Internet self-response instrument, we would observe whether the respondent provided the correct information as specified in the instructions for the vignette. We would also observe the amount of time that the respondents required to complete the Internet self-response instrument for each vignette.

After a respondent has completed all of the vignettes, we would administer a brief satisfaction questionnaire. In prior studies we have used the Software Usability Scale (<http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>) because it is very easy to administer. We have also used the Fidelity Investments Usability Scale (<http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf>) because it covers a range of usability topics.

We will debrief the respondents after they have completed the satisfaction questionnaires. The interviewers will follow a protocol, to ensure that the debriefings are uniform across the respondents. The interviewers will probe the respondents to explore the difficulties that they may have encountered as they carried out the vignettes. The interviewers will present the portions of the instruments that appeared to cause the respondents to hesitate or become confused, and ask the respondents to explain what they experienced during those times. This procedure is similar to a retrospective think aloud.

In this way, we will obtain data about the three components of usability in the NIST standards: the users’ accuracy, speed, and satisfaction in the specific context of data collection for Census 2020. Our findings will suggest the specific ways that the Census 2020 instruments might be redesigned.

### **FINAL Version**

The objectives that will guide the development of the content of the Part 1, Part 2a, and Part 2b protocols are summarized below. We will first conduct an expert review of both the self-administered Internet instrument and the NRFU instrument to identify potential usability issues. We will incorporate points to cover about these potential issues into the protocols for Parts 1, 2a, and 2b.

**Part 1: Internet Self-Response Instrument with BYOD.** Completing an Internet-based self-administered instrument is a special case of problem-solving, and therefore is very conducive to a think-aloud procedure. In the first few sessions of testing for each language, we recommend the think-aloud approach to learn in real time how the respondent responds to the translated questions and navigation options. The protocol will include instructions for interviewers to administer concurrent probes, if needed, to capture and clarify the reasons for any issues the participant encounters. We will also develop retrospective debriefing questions to elucidate any remaining issues, as well as to elicit the respondent's overall opinions about completing the self-response instrument and reactions to the supporting materials. These protocols will blend elements of cognitive interviewing and usability testing. Table 2-2 summarizes the key testing issues we anticipate for Part 1.

For the later interviews, we suggest a retrospective approach, with the interviewer observing and noting any hesitations, verbal comments, or questions the respondent asks out loud. We suggest a semi-structured retrospective probing approach for the debriefings (Willis, 2005), starting with open-ended probes to elicit respondents' explanations of their answers to each question (e.g., "How did you come up with your answer here?"). Protocols will contain additional probes to explore the respondents' observations and thought processes about the instruments, materials and translations. The protocol will also include the research goals for each of the sections of the protocol to help interviewers tailor unstructured probes. Finally, the protocol will contain probes about anything that was changed based on testing in other languages, for comparability purposes.

**Table 2-2. Potential testing objectives for Part 1, Internet self-response instrument with BYOD**

Main objectives of Internet self-response instrument testing	Additional testing issues Census may want to consider
<ul style="list-style-type: none"> <li>● Identify any comprehension issues, such as questions, specific phrases, or words in each translated instrument that are difficult to understand (e.g., unknown words or questions that are interpreted differently across respondents);</li> <li>● Identify the underlying causes of any response issues;</li> <li>● Identify any usability issues respondents encounter when completing the instrument on their own device (such as difficulties in accessing the instrument, difficulties in navigating among the screens to respond to questions, reacting to any error messages, correcting mistakes, and submitting their response); and</li> <li>● Assess reactions to the materials that will be used in conjunction with the instrument, such as what role these play in encouraging response to the Internet instrument.</li> </ul>	<p>Can respondents:</p> <ul style="list-style-type: none"> <li>● Download and install the app successfully without technical support, and remove the app at the end of the test?</li> <li>● Successfully complete the 2 factor authentication process?</li> <li>● Use their devices in both connected and disconnected mode?</li> <li>● Display multiple languages in an easily readable format and change languages while completing the instrument?</li> <li>● Navigate back to prior questions and retain previously reported responses, then move forward again?</li> <li>● Find, make use of, and understand instructions, context sensitive help, and different question types?</li> <li>● Appropriately use entry functions such as rosters, date, time, and pick lists?</li> <li>● Exit the survey prior to completion and resume successfully?</li> <li>● Understand and appropriately act on any data validation and edit messages they receive?</li> <li>● Understand and appropriately use any other special functionality incorporated into the instrument such as video, recording, or location tracking, etc.?</li> </ul>

**Part 2a: NRFU – Enumerator Perspective.** The main goal of these interviews will be to identify what kinds of issues enumerators encounter when using their specific devices to access and administer the translated instrument, and whether they can successfully enter data that their respondents provide, without making errors.

To realistically gauge the enumerator’s perspective, the protocol for testing with these participants will need to provide an overview of what the enumerator’s role is, before they start to use the instrument. The protocol will include probes appropriate for a variety of devices (iOS, Android, other as specified by Census), so as to identify issues in performing the enumerator role with the different kinds of devices and operating systems likely to be used in a BYOD context.

Working in collaboration with the Census team, we may develop a set of vignettes. Some will present common situations that all enumerators need to become proficient with; others will present very difficult or unusual situations that require more decision-making skills, such as those often associated with error in past decennial Censuses.

The interviewer will role play the vignettes with respondent/enumerators, note any confusion or issues the enumerator experiences when completing the vignette, then administer the retrospective debriefing probes to identify the causes of the issues (such as whether the

translated navigation text and labels are unclear, or the response options don't fit well with the responses the role-player provides). The debriefing will be developed as explained above, with scripted open-ended probes; it will also allow for emergent probing based on any unique enumerator experiences that surface in testing. In addition, the debriefing will also ask for other ways to fine-tune the instrument from the enumerator perspective.

In the context of the 2020 Census, enumerators will use the instrument over an extended period. It will be critical to maintain their level of motivation to do a good job and collect accurate data. Their perceived efficiency and satisfaction with the instrument will factor into their motivation. Therefore, we suggest administering a short satisfaction questionnaire as part of the debriefing. In collaboration with Census, we will identify an appropriate scale to use during the debriefing, such as the Software Usability Scale (<http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>) or the Fidelity Investments Usability Scale (<http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf>); both are easy to administer.

**Part 2b: NRFU – Respondent Perspective.** The main objectives in testing the respondent perspective of the NRFU instrument are to identify any comprehension and reporting issues respondents encounter when providing their responses to an enumerator. In these interviews, interviewers will role-play the enumerator, noting any issues with the translated questions and any confusion or hesitation the respondent exhibits. The content for the retrospective debriefing will be developed as described above. Our interviewer will repeat each question back to the respondent prior to asking any probes. The debriefing will also cover reactions to the additional materials that accompany the NRFU instrument, and important questions such as how these non-English speakers feel about providing their data in a BYOD context. For instance, are they willing to provide their data to an enumerator using their own device; do they have any concerns about confidentiality and security of their personal data?

**Parts 3 and 4, Projects D and E: Paper Self-Response Instrument, Focus Groups, Materials Testing.** For the focus groups and remaining types of interviews we will follow the same general format as described above. Most importantly, we will work collaboratively with the Census to develop the research questions and associated probes; clearly communicate the key issues for exploration to the SLCs and translators; assure quality in the translation process; and ensure the interview materials accurately reflect the Census' concerns and testing goals when they are finalized. To identify usability and translation issues in the paper self-response instruments, we propose the same think aloud and retrospective debriefing approach as for the Internet instruments. For materials testing in Projects D and E, we anticipate most of the probes will be concurrent to best gauge respondents' in the moment reactions to the materials and messages. In consultation with the Census, we may also determine that some materials require retrospective debriefing, so as to prevent reactivity (contamination of subsequent reactions by previous discussion). A likely research goal of the focus groups will be exploration of how best to motivate the Puerto Rican response. This may entail discussion questions around the kinds of messaging and language that would be most persuasive for responding via each of the modes, especially the web, since it is the first and least costly.

The Westat team will also create and translate the consent form and any other materials that may be required to facilitate the cognitive interviews. The consent form will conform to Title 13 confidentiality requirements. It will provide respondents with details about the study and inform them if the interview is being observed by Census staff, will be audio or video recorded only with the respondents' permission, is voluntary and confidential, and that all responses will be reported in aggregate with no respondents being identified. Westat will submit all materials to its IRB for approval. We are prepared to provide whatever assistance the Census team may need in preparing the OMB package.