ARTICLES

The Efficiency of Conducting Concurrent Cognitive Interviewing and Usability Testing on an Interviewer-Administered Survey

Jennifer Romano Bergstrom[1], Jennifer Hunter Childs[2], Erica Olmsted-Hawala[2], Nathan Jurgenson[3]

[1] Fors Marsh Group / U.S. Census Bureau, [2] U.S. Census Bureau, [3] U.S. Census Bureau / University of Maryland

Survey Practice

Cognitive interviewing and usability testing share techniques and have similar origins in cognitive psychology. However, cognitive interviewing often precedes usability testing in survey pretesting, and each test yields different results. While it seems that administering the tests concurrently saves time and effort, there is little evidence that researchers test concurrently. In this study, we coordinated concurrent cognitive and usability testing. By concurrently testing the same survey with different participants (respondents and interviewers), issues with question wording and design were more efficiently addressed, leading to an improved questionnaire for both respondents and interviewers.

Introduction

Cognitive interviewing and usability testing share techniques and have similar origins in cognitive psychology (Couper et al. 1999; Groves et al. 2004; Willis 2005). However, cognitive interviewing often precedes usability testing in survey pretesting, and each test yields different results. While it seems that administering the tests concurrently saves time and effort, there is little evidence that researchers test concurrently (Willis 2005). We assert that conducting cognitive interviewing and usability testing together is a more efficient approach to pretesting than conducting the tests separately.

Cognitive interviewing researchers specialize in understanding and analyzing terminology in surveys from the respondent's perspective. Usability researchers specialize in the user's experience of surveys (e.g., skip patterns, design) from the interviewer's perspective in interviewer-administered surveys. Though the techniques are similar, usability researchers often lack survey methodology training, and survey methodologists often lack usability training. Thus, while each team may recommend changes from the other's discipline, their recommendations are often not informed by experience and research. Such disjointed testing can do more harm than good. For example, a wording problem identified in usability testing, in the absence of knowledge of previous cognitive interview findings or survey methodology literature, could yield alternate wording recommendations that tested poorly in previous studies.

In this study, we coordinated concurrent cognitive and usability testing. By concurrently testing the same survey with different participants (respondents and interviewers), issues with question wording and design were more efficiently addressed, leading to an improved questionnaire for both respondents and interviewers.

Method

The Census Bureau's Center for Survey Measurement Cognitive Lab and Usability Lab usually work separately; thus, similar work that could be conducted jointly is often conducted independently. For this study, researchers from the two labs worked together on the protocols, findings, and recommendations, but different participants were recruited for each test.

The cognitive interviews focused on comprehension, accuracy, and the ability to answer questions, given participants' own situations – participants were respondents. Cognitive Lab researchers administered the survey, playing the part of interviewers. Moderators asked respondents to pretend they were at home, and moderators stood in front of respondents – simulating a doorstep interview. Moderators asked respondents to report any difficulty while attempting to answer the questions. Otherwise, interviews proceeded without interruption. Following interviews, a retrospective debriefing interview occurred, in which the moderator reviewed each of the previously administered survey questions and responses and probed about the meaning of terms and questions. Cognitive interviews examined accuracy at each of the four stages of survey response – comprehension, retrieval, judgment, and response.

Usability sessions focused on interviewers' experiences – participants were interviewers. Each session began with a brief training session. A trainer read an abbreviated training script, taken from the actual interviewer training manual. Then participants completed two practice exercises, and the trainer corrected participants as necessary, as would be done in actual training. Next participants conducted four interviews. Participants stood facing a "respondent" (Usability Lab staff) in a small room; tape on the floor marked where to stand, and participants held a notebook binder, the survey, and supporting materials. Participants asked the questions and recorded the answers. Respondents used pre-scripted scenarios that were based on prior experience with the questions (Childs 2008; Olmsted 2004; Olmsted and Hourcade 2005; Olmsted, Hourcade, and Abdalla 2005). For example, prior research (Childs 2008) revealed that scenarios with unrelated housemates are problematic for interviewers, so one of the pre-scripted scenarios was an unrelated household. After completing each scenario, participants answered retrospective probes about that scenario. Following all scenarios, participants completed a satisfaction questionnaire about the survey. Then a Usability Lab researcher (who had observed the session through one-way glass) asked participants debriefing questions about their experiences with the survey. Usability sessions measured accuracy (whether participants could complete each question per scenario), satisfaction (self-rated satisfaction ratings about using the questionnaire) and efficiency (minutes to complete each interview).
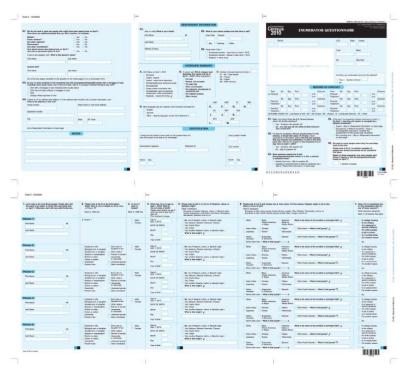
Figure 1 Front (top panel) and back (bottom panel) of NRFU survey.

Materials

The Nonresponse Followup (NRFU) survey (Figure 1) was designed to gather basic housing and demographic information for households that do not mail in the Census, and the Information Sheet (Figure 2) supplies supplementary information.

Findings[1]

Information Sheet

In the cognitive interview, respondents understood the Information Sheet and were able to use it for all relevant questions. In the usability sessions, interviewers properly administered the Information Sheet and read the associated questions. Based on results from both tests, we found that the information sheet worked well, overall. We still had several recommendations that focused specifically on interviewer training. For example, in cognitive interviewing, respondents reported that there was not enough time to read List A. Consequently, we recommended adding a note to training that the interviewer should allow the respondent time to read List A after the Information Sheet is handed to him/her.

---

1 We report notable findings. Please consult the full reports for complete findings from the cognitive interviews (Childs et al. 2009) and usability test (Romano et al. 2008).
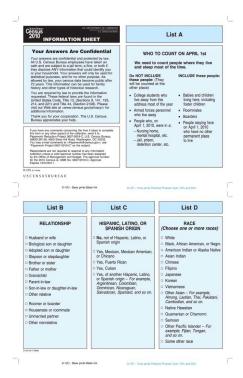
Figure 2   Front (top panel) and back (bottom panel) of NRFU Information Sheet.

Relationship Question

Joint testing allowed us to see how respondents answered questions about their households and how interviewers dealt with responses that did not map onto the response categories. For example, the Census question on relationship uses a list on the Information Sheet to display all 13 response categories to the respondent (Figures 1 and 2). The question was: "Please look at List B on the Information sheet. How is (NAME) related to (PERSON 1)?" In the cognitive interviews, many respondents reported "son" or "daughter," and the interviewer probed about whether the child was biological, adopted, or a stepchild. Some respondents could not find response categories that matched household members (e.g., a daughter's fiancé, a common law partner, a foster child). Although most respondents chose a category, it was not always the correct category that the Census Bureau uses.

In the usability tests, participants were given three test scenarios. We calculated accuracy based on how the Census Bureau would categorize relationships. Respondents were better able to classify a nanny as an "other nonrelative" (68 percent of the time) than a foster daughter (who fits into the same category and was only accurately classified 21 percent of the time). The half-sister relationship was also often misclassified (only accurately classified in 37 percent of cases).

Based on problems identified through joint testing, we recommended that training should reinforce that interviewers must ask respondents to pick the most appropriate category from List B and should then probe for "biological, adopted or stepson/daughter." We recommended that training also include examples like those that were problematic in testing (e.g., foster child).

Hispanic Origin Question

The Hispanic origin question asks respondents to look at the Information Sheet (List C, Figure 1) and answer: "Are you of Hispanic, Latino, or Spanish origin?" In cognitive interviews, some Hispanic respondents answered with "Latino" or "Spanish" without providing their countries of origin (as List B suggests). Other respondents described both their race and their origin, and some were unable to respond unassisted. With interviewer probing in the cognitive interviews, respondents were able to provide an answer that fit the Census Bureau's expectations. In the usability sessions, the test scenarios provided several different responses that respondents often give to this question, and the participants (the interviewers) were remarkably accurate (95 percent–100 percent across scenarios) in their reporting of respondents' answers.

Because respondents had difficulty in the cognitive interviews but participants playing the role of interviewers in usability testing were able to successfully navigate the question, we focused on interviewer training, which covers how to deal with respondents' problems. As usability testing suggested that the training was sufficient in this area, we had only very minor recommendations for improvement.

Conclusion

Joint testing allowed us to determine that the Information Sheet worked for both interviewers and respondents. If we had only conducted usability testing, we would not have realized that respondents merely needed additional time to process the information on the sheet. If we had only conducted cognitive interviews, we may not have known if the new format worked well for interviewers. By developing recommendations jointly, we were able to use information from both tests to help solve the respondent problem of not having enough time to process information.

For the Relationship Question, we observed where respondents had difficulty mapping their household onto response categories and where interviewers had difficulty making that assessment. The combination of these problems allowed us to conclude that interviewer training should include problematic examples and directions for probes.

For the Hispanic Origin Question, we demonstrated that respondents had difficulty with a concept, and interviewers were able to deal with that difficulty and accurately record the response. Thus, the presence of the interviewer could solve some of the problems that respondents have with this question. If we

had only conducted cognitive interviewing, we might have made significant changes to the question to assist respondents, which could have gone into the field untested because of the fast-approaching deadline. Conducting these tests concurrently allowed us to see the entire picture, from both the interviewer's and the respondent's perspectives, before recommending any changes.

Concurrent cognitive and usability testing afforded benefits beyond those provided by traditional separate testing. We gained a better understanding of how respondents and interviewers react to questions, and how interviewers react to real respondent situations. We were also able to offer suggestions to improve interviewer training based on how respondents reacted to these questions in cognitive interviews and how interviewers dealt with these issues in usability testing. If we had conducted these tests separately, suboptimal recommendations would have been generated, as the cognitive researchers would have focused mainly on the respondent, and the usability researchers would have focused on the interviewer. While we did not have time to empirically test with control groups, earlier versions of the NFRU survey underwent separate usability and cognitive tests (Childs 2008; Olmsted and Hourcade 2005; Olmsted, Hourcade, and Abdalla 2005), and findings were inefficient: The usability findings primarily focused on issues with data entry, navigation and interviewer tasks and mentioned some instances of problematic question wording; the wording problems were later replicated in cognitive interviewing.

Although the study had a very limited time for testing, the combined resources of the Cognitive and Usability Labs demonstrate how a joint process could work. Based on the quality and quantity of data gathered, we recommend concurrent cognitive and usability testing on early versions of surveys. The Census Bureau has continued to conduct combined usability and cognitive testing on a number of projects (e.g., a joint cognitive/usability test of the Puerto Rico Community Survey, see Leeman, Fond, and Ashenfelter 2012; a usability test of the 2012 National Census Test with extended debriefing and vignettes to get at cognitive aspects of user interpretation of residence rules, and targeted screens, see K. Ashenfelter et al. 2013). More combined projects are planned, but we have noticed that without conscious decisions and planning, the old practice of conducting cognitive testing separately from usability testing continues to occur. However, with more joint studies planned, and the clients' interests to streamline the pretesting process, we see this process occurring more easily and frequently into the future. As with successful website development (Romano Bergstrom, Olmsted-Hawala, Chen, et al. 2011), iterative testing should begin early in the process, with any type of survey – self- or interviewer-administered and computerized or paper (e.g., K. T. Ashenfelter et al. 2011), when time is not a constraint, and continue until optimal usability is achieved (Bailey 1993; Nielsen 1993; Willis 2005).

Acknowledgements

Author Note

REFERENCES

Ashenfelter, K., T. Olmsted-Hawala, S. Lakhe, and L. Malakhoff. 2013. "Final Report of Usability Testing of the 2013 National Census Test (NCT) Online Survey: Focus on Coverage, Race Questions, Date Format." Center for Survey Measurement Study Series (Survey Methodology #2013-01. U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2013-01.pdf.

Ashenfelter, K.T., T. Holland, V. Quach, E. Nichols, and S. Lakhe. 2011. "ACS Internet 2011 Project: Report for Rounds 1 and 2 of ACS Wireframe Usability Testing and Round 1 of ACS Internet Experiment." Center for Statistical Research and Methodology Study Series (Survey Methodology #2012-01). U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2012-01.pdf.

Bailey, G. 1993. "Iterative Methodology and Designer Training in Human-Computer Interface Design." In *The INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 198–205. http://dl.acm.org/citation.cfm?id=169163.

Childs, J.H. 2008. "2010 NRFU Questionnaire Development: From the 2004 Census Test to the 2008 Dress Rehearsal." Statistical Research Division Study Series SSM 2008-05. U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2008-05.pdf.

Childs, J.H., D.R. Norris, T.J. DeMaio, L. Fernandez, M. Clifton, and M. Meyers. 2009. "Nonresponse Followup Enumerator Questionnaire Cognitive Test Findings and Recommendations." Statistical Research Division Research Report Series RSM2009-05. U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/rsm2009-05.pdf.

Couper, M.P. et al. 1999. "The Application of Cognitive Science to Computer Assisted Interviewing." In *Cognition and Survey Research*, edited by M. Sirken, D. Herrmann, S. Schechter, and et al. New York: Wiley.

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley.

Leeman, J., M. Fond, and K.T. Ashenfelter. 2012. "Cognitive and Usability Pretesting of the Online Version of the Puerto Rico Community Survey in Spanish and English." Center for Survey Measurement Study Series (Survey Methodology #2012-09). U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2012-09.pdf.

Nielsen, J. 1993. "Iterative User-Interface Design." *Computer* 26 (11): 32–41. http://www.computer.org/csdl/mags/co/1993/11/ry032-abs.html.

Olmsted, E. 2004. "Usability Study on the Use of Handheld Devices to Collect Census Data." In *Proceedings of the International Professional Communication Conference*. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1375287&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3

Olmsted, E., and J.P. Hourcade. 2005. "NRFU Round II." Quick Report Submitted to Decennial Management Division.

Olmsted, E., J.P. Hourcade, and A. Abdalla. 2005. "NRFU 2006." Round I Submitted to Decennial Management Division.

Romano Bergstrom, J.C., E.L. Olmsted-Hawala, J.M. Chen, et al. 2011. "Conducting Iterative Usability Testing on a Web Site: Challenges and Benefits." *Journal of Usability Studies* 7 (1): 9–30.

Romano, J., E. Murphy, E. Olmsted-Hawala, and J.H. Childs. 2008. "A Usability Evaluation of the Nonresponse Followup Enumerator (NRFU) Questionnaire Form." Statistical Research Division Study Series SSM 2008-10. U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2008-10.pdf.

Willis, G.B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. London: Sage.