

# Age-Related Differences in a Usability Study Measuring Accuracy, Efficiency, and User Satisfaction in Using Smartphones for Census Enumeration: Fiction or Reality?

Erica Olmsted-Hawala<sup>(✉)</sup> and Temika Holland

Center for Survey Measurement, U.S. Census Bureau, 4600 Silver Hill Road,  
Washington DC, WA 20233, USA

{Erica.L.Olmsted.Hawala, Temika.Holland}@census.gov

**Abstract.** Age-related differences were investigated in a usability study of an application developed for U.S. Census Bureau enumerators to collect survey data and automate their time and expenses. Accuracy, efficiency and satisfaction measures were collected as participants used a smartphone to answer typical tasks. Usability flaws were also identified with the application. Results indicate that in general there were no differences with task accuracy and efficiency when comparing all tasks, however when looking at individual tasks, the task that had the most usability flaws also revealed age-related differences for accuracy and efficiency – that is older adults were less accurate and took longer to complete. Surprisingly, there were age-related differences with the user satisfaction of the application such that older adults were less satisfied with the application than younger adults. Tying age-related differences to usability flaws highlights the importance of designing optimal applications for all users.

**Keywords:** Usability · Accuracy · Efficiency · Satisfaction · Age-related differences · NRFU · Census bureau

## 1 Introduction

Every ten years, the U.S. Census Bureau conducts a mandatory census of the population. Households are encouraged to self respond, either by answering survey questions on paper and mailing the survey back, or, for the 2020 Census, answering the survey online using the Internet. However, for the households that do not respond, the Census Bureau must send a census employee (enumerator) to their door so that the household has the opportunity to answer the survey questions in person. This operation is called the Non-Response Follow Up or NRFU. The NRFU operation is a massive undertaking and in preparation for the NRFU, the Census Bureau employs temporary workers as enumerators. The enumerators are diverse in age, ranging from recent college graduates to

---

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. Any views expressed on the methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.

retirees. In fact, for the 2010 Census, 13 percent of all enumerators were over the age of 65. Forty-six percent were between the ages of 40 and 65 while 41 percent were 39 years old or younger [1]. All the enumerators for the 2010 Census conducted the NRFU on paper. However, for the 2020 Census, the business plan is to use small mobile devices (e.g., smartphones) to conduct census activities. Consequently, the software application that is created for mobile devices to aid the job of the enumerator must be suitable to enumerators of differing ages and with various levels of experience in use of smartphones.

One such prototype application under development for the 2020 Census is the Census Operations Mobile Platform for Adaptive Services and Solutions (COMPASS). COMPASS serves as an enumeration platform for conducting such activities as collecting survey data, case management, location aids, security services, and new modules that included automating time and expenses. The development team had not tested the new functionality with users and were interested in obtaining usability feedback on the new features of the COMPASS. In addition, the team was interested in identifying any usability issues that might exist in the application, including case management and icon usage on the screens.

This paper presents the results of a quantitative and qualitative usability study that investigated user behavior of effectiveness, efficiency, and satisfaction [2, 3] when using the COMPASS application.

The primary goal for the study was to identify usability issues of the application. We also wanted to get a better understanding of any performance differences among older and younger adult smartphone users. We hypothesized that (1) for the simple tasks, older adults and younger adults would be equally accurate in performance. The rationale for this was that when the task is simple, both age groups will perform with few difficulties. That is, on simple tasks, both younger and older adults will be able to complete the tasks effectively. For the complex tasks we hypothesized that (2) age and experience would come into play such that younger and older adults that were highly experienced with smartphones will perform with less difficulties, while, older adults with low to moderate experience on smartphones will have more difficulties. We further hypothesized that (3) older adults would take longer to complete the tasks. The rationale for this was twofold, older adults act slower due to, first, cognitive decline, e.g., Loos [4, 5] and Loos and Romano [6]; and second, the speed/accuracy trade off among older adults [7–10]. Finally we hypothesized that (4) there would be no age-related differences with respect to satisfaction. The rationale for this was that while intuitively it appears that satisfaction should be impacted by performance and efficiency, as we have seen in prior usability studies satisfaction rates have not been found to differ by age even when accuracy or efficiency scores did [11, 12].

## 2 Methods

### 2.1 Tasks

Participants in the usability study completed seven tasks using the COMPASS application. These tasks consisted of typical tasks that Census enumerators need to do to conduct census activities. Task difficulty was equivalent to what enumerators would do in the field

– and when initially constructing the tasks, they all appeared to be, in general, of simple cognitive complexity. These included activities such as listing the enumerators’ weekly work availability, entering their hours worked, and expenses such as tolls, and completing sample enumeration cases that targeted use and understanding/use of icons within the application. The test assessed users’ ability to perform tasks using the application and identified any problematic design features. See Appendix A for a list of the tasks.

## 2.2 Task Complexity

When initially planning the test, we intended tasks to be of the same complexity. While running participants through the usability study, however, it was clear that one task (e.g., Task 2) was proving more difficult for participants due to the usability flaws in the design of the application. Thus we categorized Task 2 as the most difficult and most cognitively challenging task.

## 2.3 Participants

Fourteen participants participated in the study: 7 younger (range 18–24) and 7 older adults (range 50–66). We divided the participants into two age groups purposely selecting age ranges that were far enough apart to detect age-related differences. The participants had at least one year experience using the Internet on a smartphone (e.g., iPhones or Androids) such as checking e-mail, getting mapping directions, reading the news, shopping online, using an app, etc. Nine of the participants were recruited from a database managed by the Center for Survey Measurement. These participants resided in the Washington DC metropolitan area and responded to a Craigslist online posting and/or flyers put up in local community centers. Five participants were former enumerators that lived in the Washington DC metropolitan area and had some prior experience in completing Census enumeration activities—however at the time of the study they were not federal employees. Participants were compensated \$40.00 for their participation. Participant demographics are presented in Table 1.

## 2.4 Procedure

Usability testing was conducted at the U.S. Census Bureau’s Human Factor’s and Usability Laboratory in Suitland, MD. The participant sat in a room facing a one-way mirror in front of a table that had the TOBII mobile eye tracker stand with the X2-60 eye tracker mounted on it. The participant entered the testing room and was informed about the purpose of the study and the use of data to be collected. The participant then signed a consent form giving permission to be audio and video recorded. The participant completed an electronic initial questionnaire about his/her smartphone use, and demographic characteristics. After that, we calibrated the participants’ eyes for eye-tracking purposes. The participant did a practice think-aloud task (e.g., the number of windows in their home) and then worked on the tasks. During the session, minimal concurrent think-aloud probing by the test administrator occurred, including such probes as “keep talking,” and “um-hum?” After the tasks, the participant answered a

**Table 1.** Mean (and range) demographics by age group

	Age group	
	Younger	Older
N	7	7
Gender	2M / 5F	3M / 4F
Age	21 (18-24)	56 (50-66)
Education	2 BA/BS	4 BA/BS
Hours per week using smartphone to access Internet?	17 (10-40)	12 (3-24)
Experience with your smartphone to use the Internet <sup>a</sup>	4.8 (4-5)	3.8 (3-5)

<sup>a</sup> Scale: 1 (None) – 5 (A great deal)

short satisfaction questionnaire to assess his/her experience using the application. Finally, we asked the participant debriefing questions about the screens and tasks that he/she had just worked on. During the session, the test administrator sat next to the participant. There were two reasons for this (1) due to the application still being in development, it could freeze up and the test administrator had to reset it. (2) The test administrator, when necessary, re-directed the participant when he/she required knowledge that he/she would learn in training, (e.g., during one task, participants needed to know that when conducting an interview with a neighbor, it was considered a “proxy visit”).

## 2.5 Usability Metrics

We assessed three typical usability metrics: accuracy, efficiency, and satisfaction. Accuracy outcomes were assigned by the test administrator and were recorded as a success (1), a fail (0), or a partial correct (0.5).

Efficiency was calculated as the total duration of the task, starting after the participant read the task aloud and ending once the participant found the answer or said they were ready to move onto the next task.

Satisfaction was calculated by summing nine scores from the modified version of the QUIS [13] administered at the end of the session. Each score was on a Likert scale from 1 to 7; so the summed score for a participant ranged from 9 to 63. The higher the score, the more satisfied the user reported being with the site.

## 2.6 Analysis Methods

Due to our small sample size ( $N = 14$ ) for accuracy we used the Fisher Exact Test with the Freeman-Halton extension in order to obtain a distribution of values in a  $2 \times 3$  table (accuracy outcome was a categorical variable with three outcomes). Using this statistic, we can decide whether the population distributions are identical. To compare differences between the two age groups in both efficiency and satisfaction we used the Mann-Whitney Test because (a) small sample size and (b) we assumed the data to be continuous but not necessarily normally distributed.

## 3 Results

We examined the relationship between age and accuracy using the Fisher exact test. Across all tasks younger adults in general performed at a higher accuracy rate than older adults. By means of the Fisher Freeman-Halton test for our  $2 \times 3$  table the relation between age and accuracy was significant  $p = 0.01$ . However when we tested each task individually, there appeared to be only one task that was making the significant difference. Task 1  $p = 0.71$ , Task 2  $p = 0.01$ , Task 3  $p = 0.56$ , Task 4  $p = 1.0$ , Task 5  $p = 0.23$ , Task 6  $p = 0.71$ , Task 7  $p = 1.0$ . With Task 2 we see that younger adults were more accurate in task performance than older adults. This task was also the most difficult for participants to accomplish due to the usability flaws in the design. Consequently when we re-run the data removing Task 2, the results for all the other tasks were not significant,  $p = 0.12$ . This indicates that for tasks that are of low cognitive complexity, with fewer usability flaws, there appear to be no age-related differences, while for the task that required more cognitive fluency, and had more usability violations, age-related differences are apparent.

We examined efficiency and satisfaction using the Mann-Whitney Test. In the descriptions below *Med* stand for Median. For efficiency, across all tasks, again while younger adults generally performed faster (in seconds) (*Med* = 168), Range (47–240) than older adults (*Med* = 334) range (68–496), like in the accuracy scores, there appear to be significant differences when looking at average time spent on all tasks  $Z = 1.85$ ,  $p \leq 0.05$ . However, as with the accuracy score, Task 2 was driving these results. When we look at the tasks individually, there were age-related differences only for Task 2, the task that was most difficult to accomplish such that younger adults were faster (*Med* = 168) range (99–224) at completing the task than older adults (*Med* = 496) range (381–660). The result is significant  $Z = 3.10$ ,  $p \leq 0.001$ . If we look at all tasks together, but remove the results from Task 2, there were no statistically significant differences between the age groups with respect to efficiency, though the trend is leaning towards significance  $Z = 1.60$ ,  $p = 0.05$ .

For satisfaction, young adult participants reported being more satisfied (*Med* = 40, range (36–45) with the application on the smart phone than their older adult counterparts (*Med* = 30), range (27–33). The result is significant  $Z = 3.53$ ,  $p = 0.0004$ .

## 4 Discussion

The accuracy results support our first hypothesis, that for simple tasks, aside from Task 2, older and younger adults do not perform significantly different from each other. Simple tasks, such as syncing the device, work for all users. The sync task which requires users to press on a visible and somewhat universal refresh symbol is not complicated such that all users in our sample, even those who use the phones less frequently are familiar with such a symbol after even a brief exposure to smartphones, and consequently are able to accomplish this with ease. Finding no age-related differences on simple tasks is also seen elsewhere in the literature (see also Olmsted-Hawala, Romano Bergstrom, Rogers [14]).

As is the case with accuracy results, efficiency results are in parallel. That is for the simple tasks, there are no age-related differences among older and younger adults when working on simple tasks. It is only with the most difficult task (e.g., Task 2) that age-related differences emerge with respect to efficiency. This is in contrast to our third hypothesis that for *all* tasks, efficiency scores differ, such that older adults take longer. This is also in contrast to the literature (e.g., 4–10) and warrants more investigation on the correlation between age and task complexity on efficiency measures. However as the *p*-value for efficiency overall approaches significance, the trend in this direction indicates the possibility that with a higher sample size we would see differences. As has been described elsewhere in the literature (e.g., Fukuda and Bubb [15]) we too find that older adults are more vulnerable to usability flaws— such that on the most difficult task where the user interface didn't meet user expectations, older adults take longer and have more difficulties in progressing successfully on the task. This is not the case for younger adults who are able to recover when confronted with the less optimal design. The complexity of the cognitive demands, in the end influences the speed with which older adults are able to accomplish their task. This is consistent with the literature (Bashore, Ridderinkhof, Molen [16]).

The ability to self-correct – e.g., make a mistake, realize it, then back up and correct the mistake by going down the more optimal path – is crucial when working on more complex tasks. When this occurred, if the participant was a young adult, they were able to self-correct, while the older adult took longer to realize, or never realized that they were in the wrong place to accomplish the task.

The issues with usability flaws and the impact associated with one's age is important for the design team and developers to take into consideration as they decide what usability fixes to make and what will be postponed or put off until the next development cycle. This is particularly important for applications that need to be optimized for adults of varying ages.

With respect to the satisfaction results, we are surprised to find that age-related differences do emerge. This is in contrast to the literature on usability study satisfaction of participants' using websites (Romano Bergstrom, Olmsted-Hawala, Jans [11]) and is contrary to our fourth hypothesis. It is interesting that older adults reported less satisfaction with the use of the application when using it on a smartphone. We speculate that the use of a small screen compounded the frustration level such that satisfaction

differences emerged. Subjective satisfaction measures with respect to age and small screens should be tested further.

#### 4.1 Limitations

A caveat to these results is the small number of participants in each age group. While usability studies in general have smaller sample sizes, typically recruiting 5 to 8, users [17, 18] our small sample does limit the statistical analyses and generalizations we can make of the data.

In terms of experience, we were unable to recruit older adults that had equal experience with the younger adults on the use of smartphones. While we did have older adults that used smartphones, they did not use them to the same extent as their younger adult counterparts. Thus it is difficult to tease out whether older adults with the same amount of experience would also have performed well on the more complex task (see also Loos [4, 5] and Hill, Dickinson, Arnott, Gregor, McIver [19]). Hence in this study, we were unable to test our second hypothesis, due to insufficient data.

It will be interesting to continue the study with additional older and younger adults and see if the trends we find hold. In addition, it would be beneficial to have more tasks of greater complexity as well as additional older adults with greater expertise in use of smartphones.

## Appendix A

Task 1: Your availability to work is as follows:

Wednesday (11/26) 8am to 4:30 pm for 6 h

Thursday (11/27) Unavailable

Friday (11/28) 8am to 12 pm for 3 h

Saturday (11/29) 8am to 12 pm for 3 h

Sunday (11/30) 8am to 12 pm for 3 h

Enter this information into the application.

Task 2: On last Tuesday 11/18 you ended up working from 9am to 12 pm. You travelled to an apartment in your assignment using your own car. You drove 13 miles to visit the apartment and 13 miles back home. You also crossed the Census Bridge, which has a toll of \$3.50 each way. Enter this information into the application and submit when complete.

Task 3: You need to do a manual sync of the data. How would you manually synchronize the data in the application? Are there any clues as to how to do this? If the data had actually synchronized, how would you be able to verify that the sync completed successfully? Tell the test administrator how you would know.

Task 4: You arrive at the first address/home you've been assigned (first one on the list). The respondent reluctantly agrees to a quick interview.

*Please begin interview and input the following responses for each screen that you encounter (in order): Personal Visit, Attempting address, Yes, Yes, Yes, Bob Terry David, 555-234-5678, No, No.*

You have arrived at the foster children screen and the respondent does not understand the purpose of this question. He asks you what the purpose is of this question. Find the answer to this question within the application.

Task 5: After you explain it to him, he suddenly grows agitated and abruptly ends the interview and refuses to answer any more questions. He goes on to say what a waste of taxpayer dollars the Census represents. No notice of visit is left as respondent orders you off his property. Exit the interview within the application and answer the questions that follow.

Task 6: You approach the second address/home of the day (second one on the list). When you arrive, you notice that the house is under construction. Based on this information, you are curious about the Contact History and Case Notes for this address. Find the Contact History and Case Notes for this house within the application.

Task 7: You are still at the same house as in Task #6 and you are growing more convinced that perhaps no one lives at this address but it's hard to tell. You then see a neighbor pull into her driveway next door. You walk over to this woman (Tammy Janice Hartmann, Phone number 202-555-5555, 345 ABC Road, Suitland, MD 20752) and ask her if she could answer a question or two about the house next door. She agrees to answer a few questions and says that no one has lived there in over 6 months (which includes July 1, 2014). She mentions that the owners abandoned the home after going way under water on their mortgage and that the bank is in the process of selling the property. She says she gets home from work around 5 pm each weekday and that it would be ok to call her if we have additional questions. Enter all of this information (including her name, phone number, address and availability) in the application.

## References

1. US Census Bureau: Internal Census Report on Age Range of Enumerators. Field Division. (2010)
2. Frøkjaer, E., Herzum, M., Hornbaek, K.: Measuring usability: are effectiveness, efficiency, and satisfaction correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, The Haag (2000)
3. Johnson, R., Kent, S.: Designing universal access: web application for the elderly and disabled. *Cogn. Tech. Work* **9**, 209–218 (2007)
4. Loos, E.F., Mante-Meijer, E.A.: Navigatie van ouderen en jongeren in beeld. Explorierend onderzoek naar de rol van leeftijd voor het informatiezoekgedrag van websitegebruikers [Older and younger adults' navigation: Explorative study on the role of age for website users' information search behaviour], Den Haag, Boom/Lemma (2009)
5. Loos, E.: In search of information on websites: a question of age? In: Stephanidis, C. (ed.) *Universal Access in HCI, Part II, HCII 2011*. LNCS, vol. 6766, pp. 196–204. Springer, Heidelberg (2011)



6. Loos, E.F., Bergstrom, J.R.: Older adults. In: Bergstrom, J.R., Schall, A.J. (eds.) *Eye Tracking in User Experience Design*, pp. 313–329. Elsevier, Amsterdam (2014)
7. Brébion, G.: Language processing, slowing, and speed/accuracy trade-off in the elderly. *Exp. Aging Res.* **27**(2), 137–150 (2001)
8. Howard, J.H., Howard, D.V., Dennis, N.A., Yankovich, H.: Event timing and age deficits in higher-order sequence learning. *Aging, Neuropsychol. Cogn.* **14**(6), 647–668 (2007)
9. Rabbitt, P.: How old and young subjects monitor and control responses for accuracy and speed. *Br. J. Psychol.* **70**, 305–311 (1979)
10. Salthouse, T.: Adult age and the speed–accuracy trade-off. *Ergonomics* **22**(7), 811–821 (1979)
11. Bergstrom, J.R., Olmsted-Hawala, E., Jans, M.: Eye tracking and Web site usability in older adults: age-related differences in eye tracking and usability performance: web site usability for older adults. *Int. J. Hum. Comput. Interact.* **29**(8), 541–548 (2013)
12. Olmsted-Hawala, E., Bergstrom, J.R.: Think-aloud protocols: does age make a difference? In: *Proceedings of Society for Technical Communication (STC) Summit, Chicago, IL* (2012)
13. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human–computer interface. In: *Proceedings of SIGCHI 1988*, pp. 213–218 (1988)
14. Olmsted-Hawala, E., Bergstrom, J.R., Rogers, W.A.: Age-related differences in search strategy and performance when using a data-rich web site. In: Stephanidis, C., Antona, M. (eds.) *UAHCI 2013, Part II. LNCS*, vol. 8010, pp. 201–210. Springer, Heidelberg (2013)
15. Fukuda, R., Bubb, H.: Eye tracking study on web-use: comparison between younger and elderly users in case of search task with electronic timetable service. *PsychNology J.* **1**(3), 202–288 (2003)
16. Bashore, T.R., Ridderinkhof, K.R., Molen, M.W.V.D.: The decline of cognitive processing speed in old age. *Curr. Dir. Psychol. Sci.* **6**(6), 163–169 (1997)
17. Nielsen, J.: Estimating the number of subjects needed for a thinking aloud test. *Int. J. Hum. Comput. Stud.* **41**, 385–397 (1994)
18. Nielsen, J., Landauer, T.K.: A mathematical model of the finding of usability problems. In: *Proceedings of ACM INTERCHI 1993*, pp. 206–213 (1993)
19. Hill, R.L., Dickinson, A., Arnott, J.L., Gregor, P., McIver, L.: Older web users’ eye movements: experience counts. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM*, pp. 1151–1160 (2011)