RESEARCH REPORT SERIES
(Survey Methodology #2020-03)

Usability of a Crowdsourcing Survey Instrument for Identifying Vacant Homes

Shelley Feuer
Erica Olmsted-Hawala
Elizabeth Nichols
Jonathan Katz
Lin Wang

Center for Behavioral Science Methods
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report issued: September, 2020

Abstract

The 2010 Nonresponse Followup (NRFU) operation was the most expensive operation in the 2010 Census and is therefore a primary focus of cost-saving research. One proposal to reduce NRFU costs is through the use of crowdsourcing - information contributed by the general public via the internet - to help identify vacant housing units sooner in the Census and survey lifecycle. This paper presents the findings of a usability experiment that assessed which type of address display (satellite map versus road map, each with an adjacent address list) allowed participants to more accurately identify vacant housing units, as well as how many residential units to include on the display. Participants were randomly assigned to use either a satellite or a road map that had either 10 or 15 neighboring housing units displayed. Using a mixed methods approach, results show that participants overwhelmingly preferred the satellite map to the road map and those who used the satellite map reported less difficulty identifying vacant units in their neighborhood and were more confident in their accuracy. Participants also preferred the maps to display 10 units rather than 15, and those who saw 10 also reported greater confidence in their task-based accuracy. Measures of eye-tracking gaze patterns corroborate participants' spontaneous and probed comments about using both the map and the address list next to the map in order to identify housing units of interest. Together, the results suggest that a 10-unit satellite map with an adjacent address list may elicit the most accurate crowdsourced data, reducing follow-up fieldwork and saving costs.

Keywords: nonresponse follow up, crowdsourcing, vacant housing

Suggested Citation: Shelley Feuer, Erica Olmsted-Hawala, Elizabeth Nichols, Jonathan Katz, Lin Wang**. (2020). Usability of a Crowdsourcing Survey Instrument for Identifying Vacant Homes.** *Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series (Survey Methodology #2020-03).* U.S. Census Bureau. Available online at <http://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/rsm2020-03.pdf>

INTRODUCTION

1. Purpose

The broad idea behind this research was to investigate whether the costs for Nonresponse Follow-up (NRFU) fieldwork could be reduced through the use of crowdsourcing - information contributed by the general public via the internet - to help identify vacant housing units sooner in the census and survey lifecycle. With the increasing popularity of online maps, one way to reduce fieldwork is by asking the public to help report vacant units in their neighborhood using an interactive online map. If the general public could identify vacant units near them, this method could potentially be used in future censuses and surveys to save costs since fewer field workers would be needed for identifying these units.

In order to do this, a person must be able to find themselves and their neighbors on a display of some kind, e.g. a map or address list, be able to accurately identify units around them that are vacant, and be willing and able to report that information. If people are able and willing to report vacancy information about their neighbors, we need to determine the best way to ask them to report it. The first step in this research was to conduct usability testing of different mapping interfaces, such as the ones shown in Figure 1.

Figure 1. Satellite (left) and Road (right) Maps

The objectives of the present study were to (1) assess which prototype maps (satellite or road) would allow participants to most accurately identify a housing unit of interest, including vacant housing units, (2) gather feedback on the optimal number of units displayed on the map, participants' map usage behavior and map preferences, and (3) gather other design interface feedback.

2. Research Questions

Primary research questions:

- Which map type and how many units allowed participants to most accurately identify a housing unit of interest, including vacant housing units?

- Which map type did participants prefer (satellite, road, list only) and how many units did they prefer be displayed (5, 10, 15)?

Secondary research questions:

- Did participants prefer that the definition of *vacant* be displayed on the screen at all times or as a hyperlink popup?
- Did they have any issues with the definition itself?
- Did participants notice the FAQ link and did they have any issues with its content or format?
- Did participants use the toggle links to change map views?

METHODS

1. Participants

Twenty-three participants (14 females, 9 males; mean age = 44.6 (*SD* = 13.7) years) from suburban neighborhoods participated in the study. Detailed demographics are presented in Table 1. A written consent to data collection was obtained from each participant. The data collection was approved by the U.S. Office of Management and Budget. Each participant received a monetary incentive for participating in the study.

Table 1. Participant Demographics

Gender (number of participants)	
Male	9
Female	14
Age	
Mean (SD)	44.6 (13.7)
Range	20-73
Education (number of participants)	
Completed high school	1
Some college, no degree	7
Bachelor degree	9
Post graduate degree	6
Hispanic origin (number of participants)	
Yes	2
No	21
Race (number of participants)	
White	9
Black or African American	8
Asian	2
More than one race selected	4

2. Map Prototypes

There were two map types, satellite and road, and two unit number displays, 10 and 15. Each map had an address list alongside it. See Figure 2.

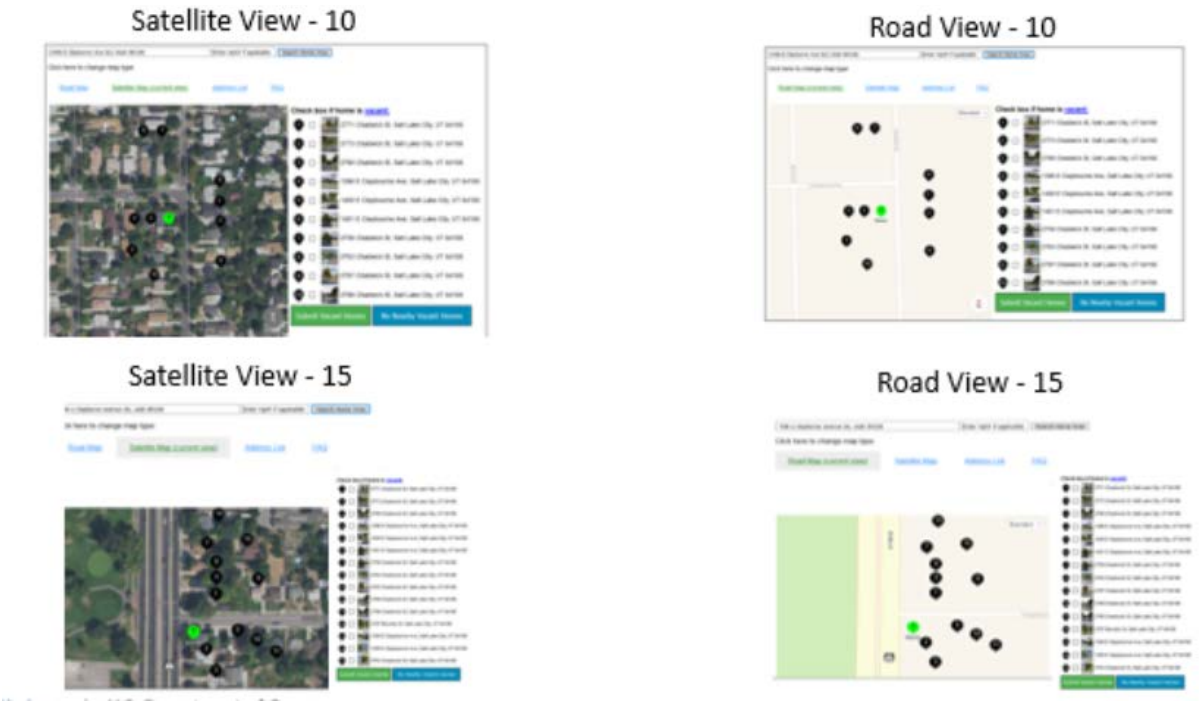

Figure 2. Map prototypes

There were also a number of other display features as shown in Figure 3. The participant's address was marked with a green pin in the map. The neighboring units (which were generated by a spiraling algorithm) were identified with other black pins. Each pin had a number that corresponded to the address list alongside the map. This list also had thumbnails (smaller images of the specific unit) next to each address. At the top of the display, there were hyperlinks (or toggle links) that allowed the user to toggle to a different map type, as well as a link for FAQ. If a user clicked on FAQ, they would receive more information about the study and how their information was protected. Participants also saw the instruction, "check box if home is vacant," and the word "vacant" was a hyperlink that, if clicked, opened a popup with the Census Bureau's definition of vacant. Finally, at the bottom, there were two buttons to submit the response: one to submit the vacant homes that they've selected and one if there were no vacant homes.

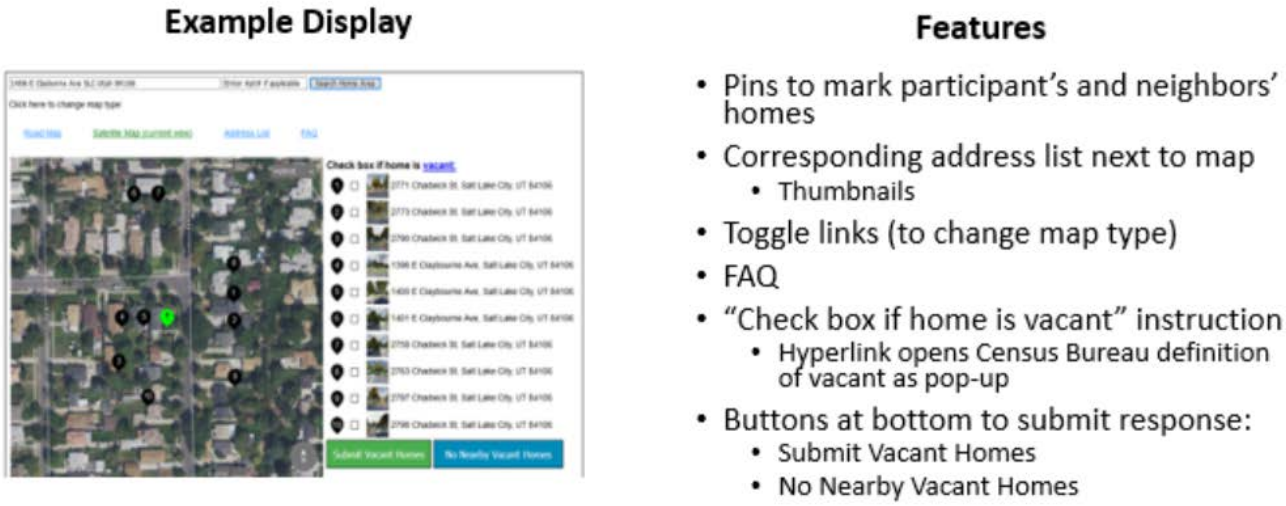

Figure 3. Example map display and list of features

3. Experimental Design

This was a between-subjects factorial design with two factors, each with two levels: map type (satellite and road) and number of units (10 and 15). This yielded four condition combinations and participants were randomly assigned to one of the four.

4. Experimental Procedure

Participants completed a demographic questionnaire and an internet-use questionnaire on the computer. We were able to collect eye-tracking data for some respondents and so for those, calibration occurred when they were seated in front of the computer. After they completed the questionnaire and they were successfully calibrated, the test administrator (TA) asked for the participant's address and typed it into the application. Participants were asked to read the instructions on the screen and were given a piece of paper with four tasks:

> Task 1: Identify and select the vacant homes in your neighborhood.
> Task 2: Identify and select the house/apartment unit that is two units away from you. Now imagine that house is vacant. What would you do?
> Task 3: Identify the homes with three or more people living there. Now imagine those same homes are vacant. What would you do?
> Task 4 (if applicable): Imagine that there are no vacancies in your neighborhood. What would you do?

Participants completed the tasks while thinking-aloud. After each task, participants completed a certainty-in-accuracy and satisfaction questionnaire on paper. Once they had finished the tasks, the TA asked a series of debriefing questions and then presented the participant with images of

each map type and unit-number display. Participants were asked which map type and how many units they preferred for finding their home and their neighbors' homes.[1]

The study followed an approach of iterative testing for usability during software development, described by Nielsen in 2000, where after about half of the sessions, the usability team identified issues that participants were struggling with, and provided design modifications to the sponsor and software developer. Once those updates were made the other half of the participants were run.

5. Performance/Outcome Measures

a. *Cognitive-behavioral feedback*: Behavioral observations, spontaneous verbalizations during live session, and verbal commentaries during debriefing
b. *Effectiveness:* Success or failure of selecting a housing unit specified in each task and number of attempts to select unit(s) of interest
c. *Satisfaction:* Participant's ratings of task difficulty on a five-point scale (1 = very easy, 5 = very difficult)
d. *Certainty-in-accuracy*: participant's rating of confidence in their task-based accuracy on a five-point scale (1 = very certain, 5 = very uncertain)
e. *Preference:* Participants' overall preference among each of the map prototypes (satellite, road, list only) and number of units displayed (5, 10, 15)
f. *Eye-tracking data:* Heat maps and other metrics (e.g. fixation counts, duration)

6. Data Analysis

Quantitative

Effectiveness: There was no validating information available so we could not determine task success (e.g., whether participants accurately selected vacant homes or homes with 3 or more people living in them) with a purely quantitative approach.[2] We analyzed effectiveness using qualitative data, whether their interaction with the application matched their verbalization and requirements of the task (e.g. clicking 'no vacant homes') and certainty in accuracy (for a more quantitative, albeit self-report, measure).

***Satisfaction and Certainty-in-Accuracy*:** We analyzed satisfaction (difficulty) and certainty-in-accuracy measures using ANOVA to determine whether condition affected self-reported task difficulty and confidence in response accuracy.

Preference: We analyzed the frequencies of map and unit number preference and compared them by condition using chi-squared tests.

[1] While the experimental design only included 10 or 15 units, after participants completed the tasks, they were also shown an example of a 5-unit display to gather feedback on the optimal number of units (in addition to 10 and 15).
[2] For task 2, however, if the participant clearly marked a home that was more or less than two units away from theirs, then that was considered a task failure

Eye-tracking: With regard to eye-tracking, there were several limitations due both to the nature of the application and the length of the interview. Consequently, eye-tracking data, including heat maps and fixation data, were only captured for five participants via a screen recording.

Application issues/limitations: Because of the computer programming used to determine which addresses to show on the map, sometimes a home right next to the participant's would not show up on the map, but homes several streets away would. The thumbnails (smaller images of the specific unit) were also missing for new housing developments. The road map was quite bare, with no other outlines or markers, like buildings or landmarks (and likely not what a final product would look like if fielded).

Qualitative: We reviewed behavioral observations and spontaneous verbalizations to identify task performance problems and reviewed any additional feedback on map designs; we reviewed probed comments from debriefing about comfort using maps and map preferences, as well as about the *vacancy* hyperlink and definition, and the FAQs.

RESULTS

Quantitative

Effectiveness: With regards to number of attempts to complete each task, most participants did not take more than one attempt to select the housing unit(s) of interest. However, those who used the road map expressed more difficulty during the think-aloud. When trying to identify and select vacant homes, one participant said, "I know my block visually but if I can't fully see it then this could be anyone's home. I can give a clear example across the street. It's vacant and has been for 2 years. But I don't know the number so this map would be confusing." There was one clear instance of task failure in the road condition. When asked to select the home two away from their own, one participant selected the one next door. He had also commented during Task 1 that he "couldn't tell which home was which home from the map."

Satisfaction and Certainty-in-Accuracy: Overall, participants rated the tasks as easy and were certain in their accuracy; lower scores indicate higher satisfaction and certainty. See Table 2.

Table 2: Mean (SD) satisfaction and certainty-in-accuracy (CIA) ratings for each task

Task	*N*	*M (SD)*
1 Satisfaction	23	2.09 (1.31)
1 CIA	23	2.43 (1.50)
2 Satisfaction	23	2.22 (1.20)
2 CIA	23	1.74 (1.10)
3 Satisfaction	23	2.30 (1.30)
3 CIA	23	2.09 (1.34)
4 Satisfaction	12	1.33 (.65)
4 CIA	12	1.58 (1.17)

The results of an ANOVA show no difference in satisfaction ratings or certainty in accuracy ratings for any of the tasks based on condition, $p > .05$. However, because of the small number of

participants in each group, we collapsed across number of units and across map type to determine whether there is a main effect of map type or number of units on ratings or if there is an interaction. Results of these factorial ANOVAs show a main effect of map type and a main effect of unit number on some of the task ratings but no interaction. Participants in the satellite conditions rated the first task (identifying vacant units) as easier (M = 1.55, SD = .69) compared to participants in the road conditions (M = 2.58, SD = 1.56), $F(1, 21) = 4.10$, $p = .05$, $\eta_p^2 = .16$. They were also more confident in their task-based accuracy (M = 1.73, SD = .79) than road participants (M = 3.08, SD = 1.73), $F(1,21) = 5.67$, $p = .027$, $\eta_p^2 = .21$. There was no effect of map type on satisfaction or confidence for the other tasks. Participants who saw 10 units were more confident that they accurately selected the unit two away from theirs (task 2) (M = 1.27 SD = .47) than those who saw 15 units (M = 2.17 SD = 1.33), $F(1,21) = 4.63$, $p = .03$, $\eta_p^2 = .22$.

Preference: Map Type: Overall, 65% of participants preferred satellite, 22% preferred road, and 13% preferred to display the list only without any map associated with it. In order to determine whether there was a relationship between map preference and map shown (i.e. condition), we conducted a chi-squared test. The results show that map preference was independent of condition, even when collapsed across unit number. Therefore, it was not significant.

Preference: Unit Number: Forty-one percent of participants preferred to see 10 units displayed and 41% preferred to see 5 displayed, while only 22% preferred 15. In order to determine whether there was a relationship between unit number preference and map shown (i.e. condition), we conducted a chi-squared test. The results show that unit number preference was independent of condition, even when collapsed across map type. Therefore, it was not significant.

Eye-tracking: A heat map comparing gaze behavior on the map and the adjacent address list shows that participants are attending to both (see figure 4); the address list has a higher fixation count and longer duration compared to the map. The time to the first fixation is also longer for the map than for the list by over 3 seconds. Because of the aforementioned limitations, we could only examine gaze patterns on the map area of the screen without separating by map type. It is also important to note that participants needed to look at the address list in order to complete the tasks, so the gaze patterns may not be indicative of preference or general utility. However, some participants commented on the combination. “The map allowed me to spatially orient myself. And the list was… where I needed to look to figure out the house number that corresponds to each bubble.” Another participant said, “I used the list but oriented myself with the area using the map.” Those who said they used the list more (or only the list) were in the road conditions. Taken together with the qualitative data, the heat map suggests that people do use both (and prefer to have both on the screen).

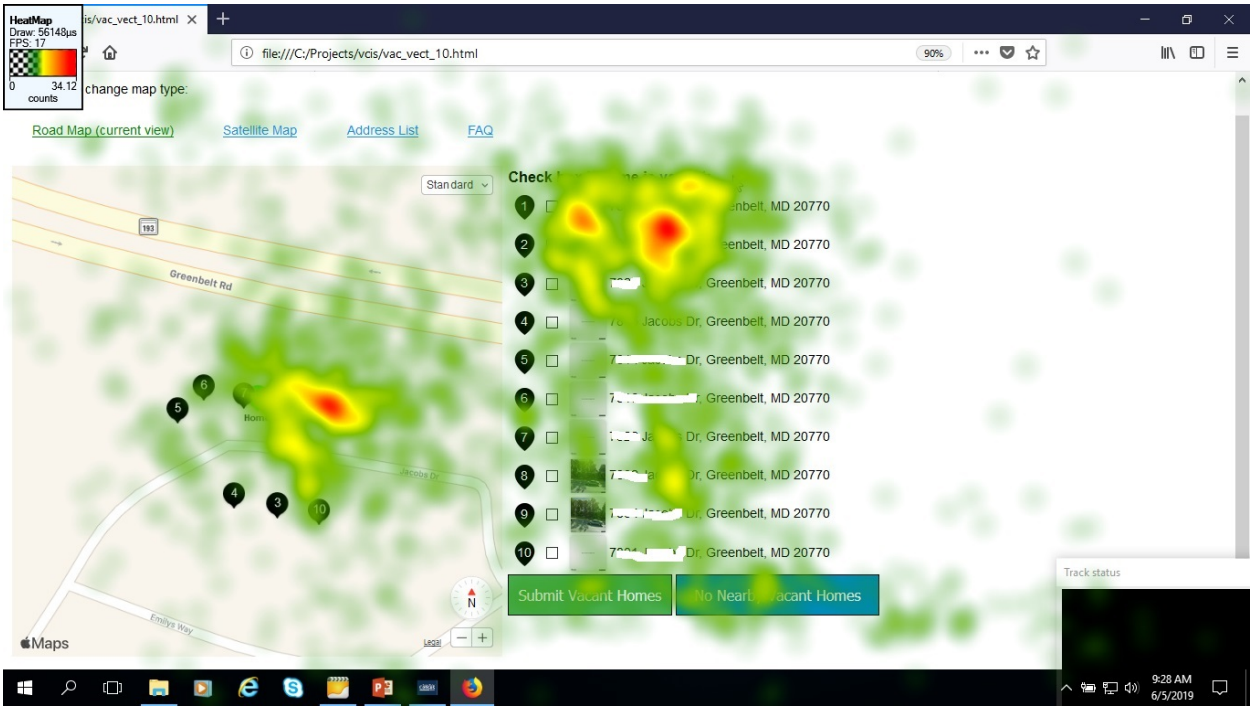

Figure 4. Heat map comparing map and address list with a road map.
Note: Figure shows a road map but heat map reflects areas of interest (AOIs), both of satellite and road

Qualitative

A few participants tried to use the (nonfunctional) zoom buttons on the map and several tried to click the icons/markers. Clicking on the icons and markers did not do anything; however we saw users attempt to click on them expecting some action to occur. One participant expected the map to interact with the address list; he said he was "hoping that clicking on the pin would highlight something on the right in the list."

Design: Some of the first half of participants tried to click on the thumbnails. After clicking on the thumbnails, nothing happened. They commented that they would like to be able to enlarge them or that they should just generally be larger. During a meeting with the developers and sponsors regarding the first round of findings, the team shared the preliminary findings and provided suggestions for design modifications. The developers changed the app so that the thumbnail could be enlarged by clicking on it. During the next half, when participants enlarged the thumbnail, a few commented that they did not know which unit it was showing and sometimes even asked for the TA's opinion. The X to close out of the enlarged thumbnail was hard to notice and participants experienced difficulty figuring out how to close it, often trying to use the backspace button or escape.

At the bottom of the address list, there were two buttons: a green "submit vacant homes" button and a blue "no nearby vacant homes" button. Some participants commented on the buttons, e.g. "blue is weird to be opposite of green. It should be a more negative color. Or maybe there should be a box at the top that says no nearby vacant homes. There should be one submit button and then

just a checkbox if there were no vacant homes. Blue is a neutral color. Not an absence color." Towards the end of the study, these findings were shared with the developers and we recommended to have only one *submit* button rather than the confusion of the two buttons and to replace the "no nearby vacant homes" button with a checkbox below the list of addresses. Due to timing, this design modification was only tested on two final participants. Both participants understood and interacted appropriately with the checkbox and the submit button.

Debriefing

Vacancy hyperlink and definition*:* The word vacant was presented as a hyperlink in the instructions, "check box if home is vacant" located above the address list. During the tasks, only a couple of participants spontaneously commented on the link and clicked on it. Most participants did not click on the link and did not notice it until the TA drew their attention to it in the debriefing.

During debriefing, some participants said that they noticed the hyperlink, though they did not necessarily realize that they could click on it. One participant said specifically "it should be clearer that it is a 'live link.'" Another participant said that she did not click the link because she knew what 'vacant' meant.

The TA asked the participants, during debriefing, to click on the link and read the definition. A couple of participants said that the definition was confusing. However, most said that it was clear and made sense, though could be shorter.

Participants were also asked if they would prefer the definition to remain as part of the hyperlink (a popup) or as text visible on the screen at all times. Most participants said that they would prefer the definition to be displayed on the screen to ensure that everyone has the same understanding of the meaning.

Toggling: There were three toggle links located at the top of the page which would open up the other map displays if the participant selected them. Only one participant toggled during the tasks. Most participants did not notice the toggle links. See Figure 4 for eye tracking heat map, which highlights that participants' eyes were not looking in that area of the screen. Participants said they were aware of the toggle link functionality when it was pointed out to them during debriefing. A couple of participants suggested that the links be placed inside the map, with images, as they are in Google maps. The usability team recommended that the developer move the three links onto the map as this was more standard formatting of mapping software and what users expected for toggling between map views. Towards the end of the cycle, the developer, unable to put them directly on the map, did place the three links as icon images below the map. This new location of the toggle links was only tested on two participants.

FAQ: Most people noticed the FAQ (when asked during debriefing) though only one clicked the link during testing. When reading the FAQs, all participants commented on the formatting, specifically, that there should be more spacing between the answers and that the questions should stand out (e.g. with bolding). "It's too dense and looks like one big paragraph." (These issues with the layout and formatting of the FAQs were identified and recommended by the usability team to

be fixed. They were addressed by the developer but again, only tested with the last two participants).

Map type and unit preference: When probed during the debriefing portion of the study, most participants responded that they would prefer the satellite map to the road map or address list. They said that the satellite view provides a clearer picture of the neighborhood. Some participants said they would prefer the road if it were more detailed.

Missing thumbnails: Participants mentioned both during debriefing and spontaneously that on some of the maps, the thumbnails were missing. Thumbnails were missing if AppleMaps did not contain an image of the front of the home. The missing thumbnails caused confusion and for some that did not have any thumbnails showing up in the view, it made the participant less sure that they were actually viewing neighborhood.

RECOMMENDATIONS

For any developers that are considering how to best have users identify their own residences or their neighbors' residences (studied here for the purpose of identifying vacant units), these usability results suggest that the satellite map should be the default map shown. Participants generally preferred the satellite map over road or list only. Those who used the satellite map reported having an easier time identifying vacant units in their neighborhood and were more confident in their accuracy. As far as number of units to display on the screen, participants equally preferred to see a list of 5 or 10; they did not prefer to see a list of 15, as the longer list seemed overwhelming and made the task seem overly cognitively challenging. Participants also used the list next to the map, according to the qualitative data as well as the eye tracking data. Having toggle links above the map were not noticed and not used by participants. Instead, toggle links to other map types should be located on the map itself as is common on other map interfaces or right below, perhaps with icons. Thumbnails in the list should definitely be larger or have enlargement capabilities, which would allow for more interaction between the list and the map. This was something that participants assumed was already a functionality as we noted many participants tried to click on the markers, or use the enlarge/minimize tools to zoom into the map. And if they can be enlarged, we observed the importance of the user being able to shrink them easily again. Finally, when at all possible, every address should have a thumbnail image associated with it. This image gave participants the assurance that they were indeed using the map correctly and seeing the property that they thought they were looking at.

The iterative nature of the usability study was effective in identifying, early on, areas of the design that did not work well for users. This approach provided the opportunity to design slight tweaks to the interface and have users interact with the changes. For such a project as this, getting real feedback from users in an ongoing process resulted in better designs overall.

Acknowledgments

We thank the Census Bureau's Center for Applied Technology for developing the experimental software used in this experiment, Mr. Sabin Lakhe for technical support, and Mr. Kevin Younes for recruitment assistance.

References

Nielsen, J. (2000). Why you only need to test with 5 users. Nielsen Norman Group. Available at: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/