Usability Evaluation Plan for CEDSCI (Phase I)

Draft: September 10, 2016
Last revision: October 31, 2016

Version 3.1

<u>Study Team</u>

HFUR/CSM and CEDSCI/Decennial

<u>Sponsors</u>

ADDCP

U.S. Census Bureau

1. Introduction

The CEDSCI is an initiative to enhance the U.S. Census Bureau's data dissemination capability, and consequently to foster a user-focused and cost-effective data dissemination environment. The goal of CEDSCI is to provide users with credible statistics, easy-to-use tools, and standardized data elements. The CEDSCI will be carried out in phases. In Phase I, an application will be developed for users to search for certain demographic and economic information collected by the Census Bureau. The application will have two versions, one for desktop and the other for smartphone. This usability evaluation project is to assess the usability of the Phase I application.

The purpose of this project is to optimize the application's usability through **iterative usability evaluations during the entire lifecycle of the product design and development**. Usability will be evaluated in terms of respondent's **effectiveness** and **efficiency** in survey completion, and **satisfaction** with the experience of interacting with the application.

2. Project Objectives

This project is to accomplish the following objectives:
1) Minimize potential usability problems during application design phase.
2) Identify and address usability problems during application development phase.
3) Assess application's usability upon the completion of development.
4) Establish performance benchmarks for the application.

3. Application Development Lifecycle

The application will be developed using the Agile methodology. From October 1, 2016 to September 30, 2017, there will be 4 major releases at a 90-day interval. Within each 90-day interval, there will be 6 sprints of 15 days each.

Due to the nature of Agile development, detailed functionalities of the application will be designed in sprints. In general, the application will have the following functional components: Navigation, data search, text display, graphical display, data download, and etc.

4. Project Scope

The usability evaluation will include **four aspects**, to be conducted in order:
1) Baseline summative testing of the existing AFF application.
2) Usability review of the high-level information architecture.
3) Rapid formative usability testing at the completion of each sprint.
4) Comprehensive formative testing at each release.
5) Summative testing at the application release.

Table 1 outlines estimated project schedule for the aforementioned tasks.

Table 1. Estimated Project Schedule

Activity	Start	End	Duration (day)
Baseline summative testing of the existing AFF application			20
Information architecture review			5
90-day Interval I			
Rapid formative test, Sprint 1			2

Rapid formative test, Sprint 2			2
Rapid formative test, Sprint 3			2
Rapid formative test, Sprint 4			2
Rapid formative test, Sprint 5			2
Rapid formative test, Sprint 6			2
Comprehensive formative testing			20
90-day Interval II			
Rapid formative test, Sprint 1			2
Rapid formative test, Sprint 2			2
Rapid formative test, Sprint 3			2
Rapid formative test, Sprint 4			2
Rapid formative test, Sprint 5			2
Rapid formative test, Sprint 6			2
Comprehensive formative testing			20
90-day Interval III			
Rapid formative test, Sprint 1			2
Rapid formative test, Sprint 2			2
Rapid formative test, Sprint 3			2
Rapid formative test, Sprint 4			2
Rapid formative test, Sprint 5			2
Rapid formative test, Sprint 6			2
Comprehensive formative testing			20
Final Release			
Final summative testing – a test of the overall product			30

5. **Baseline Summative Testing**

This summative usability test is to establish a baseline performance of the existing AFF application which is the precursor of CEDSCI. The usability information collected in this test will be used for comparison with that of the final summative testing after the formal release. During the summative testing, the participant will complete a set of tasks on his/her own without interruption or probes.

5.1. Methods

The summative testing will be conducted in the same fashion as the comprehensive formative testing except for the following. A set of tasks that involves all major functionalities of the application will be performed by the participants. The participant will perform the tasks without assistance or interruption. Think-aloud and probing during the task performance will not be conducted. Task completion time and errors will be recorded.

Outcome measures. (1) Task completion time. (2) Errors.

Test design. TBD.

Participants. 20 individuals (10 males and 10 females) will participate in the testing, desktop version only. The same inclusion criteria and recruitment strategies as that in the rapid formative testing will be applied.

Equipment. The participants will use the Census-Bureau-provided laptop/desktop computer.

Test case. TBD.

Data collection methods. Same as the formative testing, except that there will be no think aloud and debriefing may be conducted accordingly.

Data analysis. Same as the formative testing, with additional quantitative analysis.

Testing procedure. Same as the formative testing, except that there will be no think-aloud.

Testing facility. Same as the formative testing.

One session is estimated to last about 90 minutes. This task is expected to be completed in 20 work days.

6. Information architecture review

Upon completion of the information architecture design, the usability evaluation team and the sponsor team will jointly review the design from the usability perspective, and to provide suggestions of usability improvements. The purpose of this activity is to enhance a human-centered design of the application, and to minimize potential usability problems down the road.

HFUR/CSM and CEDSCI/Decennial will review whatever information is at hand and provide assessment to the current state of the information architecture. The review will pay particular attention to the specific functional goals and related roadmap or strategies to realize the goals.

6.1. Methods

The reviewers will meet and walk through the design according to the functional requirements. A review protocol may need to be established.

This task is expected to be carried out before development phase and to be completed in 5 work days.

7. Rapid formative usability test

Upon the completion of each sprint, a formative usability test will be conducted on both desktop and smartphone versions of the application, on a small sample of potential users. The purpose of the test is to discover major usability problems promptly so that the development team can address the problems in a timely fashion. During the test, the participants' performance will be investigated using a think aloud protocol.

7.1. Methods

For each rapid test, a set of test cases will be designed according to the functionalities developed in the sprint. The participant will be required to complete tasks in the test cases. During the task performance, the participant will be asked to think aloud (verbalizing what he/she is thinking) and will be probed as needed. In addition, a mouse-tracking technique will be employed to record the participant's visual scan and gaze pattern, when it is feasible. After the completion of the tasks, the participant will be debriefed about his/her experience with the application.

Participants. Six individuals will participate in the testing, 3 for desktop version and the other 3 for smartphone. Participants will be recruited primarily based on their demographic compatibility with what is required for target user population. The following approaches will be used in recruitment:

    a. Word of mouth, craigslist ads, flyers posted in libraries, and the recruiting database, where appropriate
    b. Sponsor's assistance with recruiting by contacting their sources
    c. Notifying the sponsor of any difficulty recruiting by the 2nd week of testing

Test design. The rapid formative test will consist of 3-5 usability tasks based on what is developed during a 2-week sprint. Each test will be conducted separately on a 2-week cycle between the baseline and subsequent formal formative testing and will identify immediate feedback for the development team.

Given the fluid nature of the Agile development process, the exact tasks for each rapid test will need to be determined by the time when it is close to the testing date. Nevertheless, a set of anticipated tasks has been identified in the test case section below.

Test case.
Home page and warning page –
Probes -
- Do users clearly understand what to do with the CEDSCI tool from the home page?
- Do users understand that they will be using a tool being built in progress?

Search

Task	Scenario
Filter by topic	You are interested in finding out what the typical commute is in your area, where would you look to find this?
Filter by geography list	You are considering moving to Accomack County because you hear the schools are particularly good, you would like to know what the average mortgage is for a house in the county, how would you find this information?
Filter by geography using the selection map	You would like to know what a typical mortgage is in a neighboring county to Accomack county, how would you do this?
Filter by industry	You would like to know how many grocery stores there are in the United States, using this tool, how would you find this information?
Filter by year	You would like to find out the poverty level for the year 2013, how would you do

Task	Scenario
	this?
Filter by Dataset (Expert Users)	Determine what you would find in each dataset (open filter and go down list, ask if results were expected.)
Filter by Survey/Program	Determine what you would find in each Survey/Program (open filter and go down list, ask if results were expected.)
Filter by multiple combinations of the above	You are doing research around construction companies and would like to know how many companies existed in 2013 in Washington, DC, how would you do this?
Finding total state population	You are doing background research and would like to know how many people live in the state of California in 2013, how would you find this information?
Find total county population	You are doing background research and would like to know how many people live in the Chase County, Nebraska , how would you find this information?

Probes – Do users notice the Topic/List view?

Search Results

Task	Scenario
Filter results	You would see how many construction companies existed in 2012 in Washington, DC, how would you do this?
Select geographies from a table	You would like to see how many construction companies existed in 2012 in Virginia, how would you do this?
Map data	You would like to see a visual representation of construction companies that existed in Maryland, Virginia and Washington, DC how would you do this?
Downloading data tables	You would like to process this data table with a program on your computer, how would you obtain a file of this table?
Viewing Footnotes	As you are looking at a table you notice a "**" in the margin of error column, you would like to know what "**" means. Where would you look for this information?

Probes –

Are users confused by changing data sets?
Do users click on the title of the result or do they click on the table ID?
Do users know what a table ID is?
Do users notice Quick Facts on the side of the search results page?
Do users understand what the "Web Pages" tab is? Would they use it?
Do users see and understand footnotes?

Data collection methods. The following methods will be used to collect participants' performance data:

a. Think-aloud protocol with minimal probing such as "Keep Talking," "What are you thinking?" and acknowledgement tokens (linguists refer to this as backchannels) such as "Um-hum?"
b. Observation notes
c. Satisfaction questionnaire
d. Retrospective Debriefing
e. Audio and video recording
f. Mouse tracking recording

Data analysis.

Qualitative

a. Review behavioral observations, spontaneous verbalizations and answers to debriefing questions in order to identify problems
b. Produce gaze patterns on PC to investigate whether participants attended to or ignore important parts of the screens. We will investigate whether we can produce gaze patterns on smartphones.

Quantitative

a. Compute the overall satisfaction ratings

Testing procedure. The session will be conducted one-on-one, i.e., one participant and one test administrator (TA), with one note taker. The test will be carried out in the following sequence:

a. Participant's consent
b. TA start audio and video recording
c. Participant complete tasks, while thinking-aloud
d. Participant completed satisfaction questionnaire
   TA conduct retrospective Debriefing

Testing facility. A testing session will be carried out in either the usability lab or a community library.

One session is estimated to last about 60 minutes. This task is expected to be completed in 2 work days.

8. Comprehensive usability test

Upon the completion of each 90-day release, a comprehensive formative usability test will be conducted on both desktop and smartphone versions of the application, on a relatively large sample of potential users. The purpose of the test is to assess the overall

usability of the release and to identify usability problems that need to be addressed. During the test, the participants' performance will be investigated using a think-aloud protocol.

8.1. Methods

For each comprehensive test, a suite of test cases will be designed according to the functionalities developed for the release. The participant will be required to complete tasks in the test cases. During the task performance, the participant will be asked to think aloud and will be probed as needed. In addition, the eye-tracking technique will be employed to record the participant's visual scan and gaze pattern, when it is feasible. After the completion of the tasks, the participant will be debriefed about his/her experience with the application.

Participants. 18 individuals will participate in the testing, for desktop or smartphone versions, depending on the nature of test cases. The same inclusion criteria and recruitment strategies as that in the rapid formative testing will be applied.

Equipment. The participants will use the Census-Bureau-provided laptop/desktop computer or smartphones for the testing.

Test design. TBD
Test case. TBD
Data collection methods. Same as that of rapid formative testing.
Data analysis. Same as that of rapid formative testing.
Testing procedure. Same as that of rapid formative testing.
Testing facility. Same as that of rapid formative testing.

One session is estimated to last about 90 minutes. This task is expected to be completed in 20 work days.

9. Final Summative Testing

Before the instrument goes to production, a summative usability test may be conducted. The purpose of the summative test is to establish user performance benchmarks for major use cases. During the summative testing, the participant will complete a set of tasks on his/her own without interruption or probes.

9.1. Methods

The summative testing will be conducted in the same fashion as the comprehensive formative testing except for the following. A set of tasks that involves all major functionalities of the application will be performed by the participants. The participant will perform the tasks without assistance or interruption. Think-aloud and probing during the task performance will not be conducted. Task completion time and errors will be recorded.

Outcome measures. (1) Task completion time. (2) Errors.
Test design. TBD.
Participants. 20 males and 20 females will participate in the testing, 20 for desktop version and the other 20 for smartphone. The same inclusion criteria and recruitment strategies as that in the rapid formative testing will be applied.

Equipment. The participants will use the Census-Bureau-provided laptop/desktop computer, but their own smartphones for the testing.

Test case. TBD.

Data collection methods. Same as the formative testing, except that there will be no think aloud and debriefing may be conducted accordingly.

Data analysis. Same as the formative testing, with additional quantitative analysis.

Testing procedure. Same as the formative testing, except that there will be no think-aloud.

Testing facility. Same as the formative testing.

One session is estimated to last about 90 minutes. This task is expected to be completed in 30 work days.