Standards and Guidelines for Designing Human-Centered Mobile Surveys

Lin Wang, Elizabeth Nichols, Erica Olmsted-Hawala, Shelley Feuer, Alda Rivas
Christopher Antoun, Brian Falcone, Ivonne Figueroa

Center for Behavioral Science Methods
U.S. Census Bureau

APPOR Webinar September 15, 2022
Disclaimer

This document is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not those of the U.S. Census Bureau.
Acknowledgement

This project is supported by the U.S. Census Bureau Innovation and Operational Efficiency Program

Associate Directorate for Research and Methodology
Associate Directorate for Information Technology
Center for Behavior Science Methods
Information Systems Support and Review Office
The Center for Applied Technology
MetroStar Systems

Russell Sanders (Project Manager)

John Abowd (Champion)   Sabin Lakhe   Chin Fang Weng
Paul Beatty (CBSM Chief) Lawrence Malakhoff Luke Larson
Benjamin Saunders (COR)  Kevin Younes   Joanna Fane Lineback
Harry Lee (former Champion) Marylisa Gareau Wei Chen
Barbara LoPresti (former Advisor) Rebecca Keegan Patricia Summers
Alexandra Figueroa(former Sponsor) Jonathan Katz Josef Argueta
Thomas Mathew          Andrew Raim       Ashley Peterson
Challenges in Mobile Survey Design
Problem

- Inappropriate design of user interface
- Misperception of survey questions
- Erroneous responses

Groves et al, 2004
Solution

- Develop **standards** and **guidelines** for mobile survey design to minimize measurement errors.
- Standards concern the **basic** operations across different elements of survey responses, serving as rules.
- Guidelines concern specific **survey** response operations, serving as “best practices.”
- All standards and guidelines are based on **empirical evidences**.
A Systematic Approach

Step 1: Developing a mobile survey respondent model
Step 2: Formulating topics for standards and guidelines
Step 3: Gathering evidences
Step 4: Making standards/guidelines based on evidence
A Mobile Survey Respondent Model

Distribution of potential respondents in terms of ability

<table>
<thead>
<tr>
<th>Dimension I: Near vision (for reading)</th>
<th>Habitual visual acuity: reading newspaper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrast sensitivity: usual (newspaper print)</td>
</tr>
<tr>
<td></td>
<td>Color vision: impaired</td>
</tr>
<tr>
<td>Dimension II: Index finger</td>
<td>Operating fingertip breadth: 13 mm</td>
</tr>
<tr>
<td></td>
<td>Operating finger mobility: stiff but able to operate a smartphone</td>
</tr>
<tr>
<td>Dimension III: Cognitive ability</td>
<td>Mentally alert</td>
</tr>
<tr>
<td></td>
<td>Language: fluent in English</td>
</tr>
<tr>
<td></td>
<td>Education: 8th grade or equivalent</td>
</tr>
</tbody>
</table>
Developing Topics for Standards

- Standards concern the basic operations across different elements of survey responses, serving as rules.
- Three areas of user interface elements:
  1) Touch target
  2) Text entry and display
  3) Text/graphics luminance and color

Topic candidates:
- Touch target size
- Touch target spacing
- Text field height
- Text size
- Foreground/background luminance contrast
- Color combination
Guidelines concern specific survey response operations, serving as “best practices.”

### Developing Topics for Guidelines

**Iterative Brainstorming**

1\(^{st}\) brainstorm session: UX researchers with survey instrument testing experiences proposed a wish list of topics

Lead researcher consolidated the list

Follow-up brainstorm session: Team researchers review and revise the consolidated list

Final topic list

### Survey Question Model

- Pre-question-stem instruction
- Question stem
- Post-question-stem instruction
- Response options
- Post-response instruction

### Topic Areas

<table>
<thead>
<tr>
<th>Category</th>
<th>Sample Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question instruction</td>
<td>Layout</td>
</tr>
<tr>
<td>Question stem</td>
<td>Text color</td>
</tr>
<tr>
<td>Response</td>
<td>Response options</td>
</tr>
<tr>
<td>Response orientation</td>
<td>orientation</td>
</tr>
<tr>
<td>Navigation</td>
<td>Optimal navigation method</td>
</tr>
<tr>
<td>Support features</td>
<td>Within-question Help link</td>
</tr>
<tr>
<td>General</td>
<td>Text-Field Labeling</td>
</tr>
</tbody>
</table>
Gathering Evidences

Two-step approach:

- Step 1 – Gathering existing evidences in literature
- Step 2 – Generating evidences through behavioral experiments

Comparative outcome analysis:

Do participants <perform a task> more effectively, more efficiently, with better satisfaction <using> <this design> than <other design(s)>?

**Effectiveness:**
Accuracy and completeness with which participants perform the task.

**Efficiency:**
Resources used in relation to the results achieved.

**Satisfaction:**
The extent to which the participants’ physical, cognitive, and emotional responses that result from task performance.
## Literature Review

<table>
<thead>
<tr>
<th>Guide No</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>2</td>
</tr>
<tr>
<td>Category</td>
<td>General</td>
</tr>
<tr>
<td>Topic</td>
<td>Screen Orientation</td>
</tr>
<tr>
<td>Research questions</td>
<td>Should the survey instrument be designed for portrait or landscape display?</td>
</tr>
<tr>
<td>Guide</td>
<td>Design questionnaires optimized for portrait orientation.</td>
</tr>
<tr>
<td>Evidence</td>
<td>Paper 1 ... Paper 2 ...</td>
</tr>
<tr>
<td>Evidence strength</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

---

A three-level evidence strength rating system:

- **Strong** - Two or more peer-reviewed studies.
- **Moderate** - Single peer-reviewed study in conjunction with at least two non-peer-reviewed reports.
- **Weak** - Two or more non-peer-reviewed reports.

*A guideline must be supported by at least two studies.*
Behavioral Experiments

**Research question:**
Is Design A better than B, C ...?

**Performance measures:**
Effectiveness, Efficiency, Satisfaction.

**Experimental factors:**
Design A, B, C ...

**Experimental paradigm:**
Task design.

**Experimental design:**
Factorial design.

**Participants sample:**
30 individuals between 60 and 75 years old (target population: 18+)

---

*Fig. 2. From Schefrin et al. (1999). Average scotopic spatial contrast sensitivity for three age groups. Squares signify 20–40-year-olds, circles 41–60-year-olds, and triangles 61–88-year-olds. Note that these losses are sizeable at low spatial frequencies and cannot be attributed to optical factors, suggesting a neural origin.*
Protocol for Determining a Standard/Guideline

Do participants *perform a task* more effectively, more efficiently, with better satisfaction *using* *this design* than *other design(s)‽*

- **Is one design more effective than other designs?**
  - Yes: The more effective design is selected
  - No: Next question

- **Is one design more efficient than other designs?**
  - Yes: The more efficient design is selected
  - No: Next question

- **Is one design more satisfying than other designs?**
  - Yes: The more satisfying design is selected
  - No: All or no designs are selected
## Current Version of Standards and Guidelines for Mobile Survey Design

### Standards

<table>
<thead>
<tr>
<th>Category</th>
<th>Relevant Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touch target size</td>
<td>Standard 1</td>
</tr>
<tr>
<td>Text display</td>
<td>Standard 2</td>
</tr>
<tr>
<td>Luminance and color</td>
<td>Standard 3, 4</td>
</tr>
</tbody>
</table>

### Guidelines

<table>
<thead>
<tr>
<th>Category</th>
<th>Relevant Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire display or layout</td>
<td>Guideline 1 - 5</td>
</tr>
<tr>
<td>Supporting information display (e.g., help link)</td>
<td>Guideline 6 - 8</td>
</tr>
<tr>
<td>Login ID entry</td>
<td>Guideline 9</td>
</tr>
<tr>
<td>Navigation</td>
<td>Guideline 10</td>
</tr>
<tr>
<td>Labeling of action buttons</td>
<td>Guideline 11 - 14</td>
</tr>
<tr>
<td>Question stem and response option</td>
<td>Guideline 15 – 26</td>
</tr>
<tr>
<td>Interviewer-administered surveys</td>
<td>Guideline 27 - 30</td>
</tr>
</tbody>
</table>
Sample Standards and Guidelines

- Standard 4 – Literature review
- Standard 1 – Behavioral experiment
- Guideline 12 - Behavioral experiment
- Guideline 19 – Behavioral experiment
- Guideline 15 – Behavioral experiment
Standard 4 (I)

Use of Color: avoid placing red and green colors next to each other

Red-Green
(8% man, 0.5% women)

Blue-Yellow

good color vision
blue-yellow deficient

Total Color Blindness

92% Normal Vision
2.7% Deuteranomaly
0.66% Protanomaly
0.59% Protanopia
0.56% Deuteranopia
0.016% Tritanopia
0.016% Tritanomaly
<0.0001% Achromatopsia
Standard 4 (II)
Use of Color: avoid placing red and green colors next to each other

Evidence:
A well-established clinical fact (https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness)

Research question:
What is the optimal size of a touch target and spacing in mobile surveys?

Performance measures:
- Target touch
- Target touch time
- Task difficulty rating

Success criterion:
Target touch success rate ≥ 80%
Standard 1 (II)
Size of Touch Button: at least 6 mm of square side or circle diameter

Experimental design:
Two-factor within-subjects design – Target size, Target spacing. 75 combinations.

<table>
<thead>
<tr>
<th>Length of square side (mm)</th>
<th>Max spacing (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>
Standard 1 (III)

Size of Touch Button: at least 6 mm of square side or circle diameter

Experimental paradigm:
Touching a target on a smartphone screen.

Experimental procedure:
1. Instruction.
2. Practice.
3. Task performance.
4. Easiness rating.

Performance measure definitions:
*Touch*: Success, Failure
*Touch Time*: Time from leaving the Start button to touching the target
*Task Difficulty*: 5-point scale with 1 being “Very easy” and 5 “Very difficult”
Standard 1 (IV)
Size of Touch Button: at least 6 mm of square side or circle diameter
Standard 1 (V)
Size of Touch Button: at least 6 mm of square side or circle diameter

Participants sample:
- N = 52 (Male: 15, Female 37)
- Age (year): Mean (SD) = 68.74 (5.38), Range: 59 - 80
- Education: High school or above
- Spanish origin: Yes = 4%, No = 96%
- Race: White = 67%, Black = 13%, Asian = 13%, Other = 7%
Standard 1 (VI)

Size of Touch Button: at least 6 mm of square side or circle diameter

The relationship between target-touch success rate and target size

- Generalized estimating equation (GEE) of a binomial distribution, logit link function, and exchangeable correlation matrix.
- **Target size $\uparrow \rightarrow$ touch success rate $\uparrow$** ($p < 0.0001$)
- No significant effects of target spacing nor of the interaction between target size and target spacing.
- **Where is the change point?**

<table>
<thead>
<tr>
<th>Target Size (mm)</th>
<th>Target-touch Success Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>18.46</td>
</tr>
<tr>
<td>3</td>
<td>35.91</td>
</tr>
<tr>
<td>4</td>
<td>53.85</td>
</tr>
<tr>
<td>5</td>
<td>75.00</td>
</tr>
<tr>
<td>6</td>
<td>86.77</td>
</tr>
<tr>
<td>7</td>
<td>90.38</td>
</tr>
<tr>
<td>8</td>
<td>92.69</td>
</tr>
<tr>
<td>9</td>
<td>94.62</td>
</tr>
<tr>
<td>10</td>
<td>97.69</td>
</tr>
<tr>
<td>11</td>
<td>99.62</td>
</tr>
</tbody>
</table>
**Standard 1 (VIII)**

Size of Touch Button: at least 6 mm of square side or circle diameter

**Change point detection**
- Bayesian inference with Markov Chain Monte Carlo (MCMC) simulations.

$$y(x) \sim \begin{cases} \text{normal}(\alpha + \beta_1(x - cp), \sigma^2), & x < cp \\ \text{normal}(\alpha + \beta_2(x - cp), \sigma^2), & x \geq cp \end{cases}$$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Mean</th>
<th>STD</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>20000</td>
<td>86.4701</td>
<td>2.1036</td>
<td>82.5055 – 90.6716</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>20000</td>
<td>18.6781</td>
<td>0.9742</td>
<td>16.6761 – 20.4971</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>20000</td>
<td>2.5285</td>
<td>0.5318</td>
<td>1.4311 – 3.5431</td>
</tr>
<tr>
<td>$cp$</td>
<td>20000</td>
<td>5.6863</td>
<td>0.1912</td>
<td>5.3359 – 6.0966</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>20000</td>
<td>21.8097</td>
<td>4.5987</td>
<td>13.7218 – 30.9075</td>
</tr>
</tbody>
</table>

(CBDRB-FY22-CBSM002-042)
Standard 1 (IX)

Size of Touch Button: at least 6 mm of square side or circle diameter

The relationship between target-touch time and target size

- Only observations of successful target-touch were included in this analysis.
- The log-transformed data were fitted to a GEE model.
- **Target size ↑ → touch time ↓** \( (p < 0.0001) \)
- No significant effects of target spacing nor of the interaction between target size and target spacing.
Standard 1 (XI)

Size of Touch Button: at least 6 mm of square side or circle diameter

**Difficulty rating**
- Five-point scale with 1 being very easy and 5 being very difficult.
- Data were collected from 50 participants.
- Mean rating of 1.3 with SD of 0.54.
Guideline 12 (I)
Label a Logout button with the text of “Save and Logout”

Research question:
Should “save and logout” functionality be explicitly labeled on a button?

Experimental design:
Between-subjects design with one 2-level factor, Button Label:
- a. Logout
- b. Save & Logout
Guideline 12 (II)
Label a Logout button with the text of “Save and Logout”

Experimental paradigm:
Step 1: The participant (P) was shown a mock-up design (either “Save & Logout” or “Logout”).
Step 2: P was asked two probing questions about the design.
Step 3: Both designs were presented side-by-side to P. P was then asked to select a preferred design.

Probing questions:
Q 1: If you had to go somewhere before finishing the entire survey, what would you do?
Q 2: Do you think any information you have already entered will still be there when you return and log back into the survey later?
Guideline 12 (III)
Label a Logout button with the text of “Save and Logout”

<table>
<thead>
<tr>
<th>Performance Measure Coding Schemes</th>
</tr>
</thead>
</table>
| **Likelihood of logout button use** | Question 1:  
Coded “1” – Participant said they would select “Logout” button (in Condition 1) or “Save and Logout” button (in Condition 2)  
Coded “0” – Anything else |
| **Interpretation of button functionality** | Question 2:  
Coded “1” – Participant responded information would be saved  
Coded “0” – Participant responded information would NOT be saved |
| **Preference** | Overall preference between the two label designs |
Guideline 12 (IV)
Label a Logout button with the text of “Save and Logout”

Participants sample:

- N = 54 (Male: 20, Female 34)
- Age (year): Mean (SD) = 52.5 (20.1), Range: 21 - 85
- Education: High school or above
- Spanish origin: Yes = 9%, No = 91%
- Race: White = 61%, Black = 26%, Asian = 11%, Other = 2%
Guideline 12 (V)
Label a Logout button with the text of “Save and Logout”

Results:

Likelihood of logout button use – Participants (P) presented with “Save & Logout” button would more likely to use the button ($p = 0.003$). (Table A)

Interpretation of button functionality – Ps presented with “Save & Logout” button would more likely to believe in information being saved ($p < 0.001$). (Table B)

Preference - All Ps preferred the “Save and Logout” button over the “Logout” button.

<table>
<thead>
<tr>
<th>Table A</th>
<th>Logout button use</th>
<th>Condition (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Save &amp; Logout</td>
<td>Logout</td>
</tr>
<tr>
<td>Yes</td>
<td>88.9% (24)</td>
<td>48.1% (13)</td>
</tr>
<tr>
<td>No</td>
<td>11.1% (3)</td>
<td>51.9% (14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table B</th>
<th>Response Option</th>
<th>Condition (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Save &amp; Logout</td>
<td>Logout</td>
</tr>
<tr>
<td>Information Saved</td>
<td>96.3% (26)</td>
<td>26% (7)</td>
</tr>
<tr>
<td>Information Not Saved</td>
<td>3.7% (1)</td>
<td>74% (20)</td>
</tr>
</tbody>
</table>
Guideline 19 (I)

Use 6-mm radio-button/check-box for response options

Research questions:
1. Do large response-option icons improve respondent performance compared to small icons?
2. Do wide buttons for response options improve respondent performance compared to conventional response-option buttons?

![Small icon](image)
![Large icon](image)
![Wide button + icon](image)
![Wide button](image)
Guideline 19 (II)

Use 6-mm radio-button/check-box for response options

Experimental design:
Between-subjects design with one 4-level factor, Response Option:

a. Conventional control: 2-mm circular button
b. Larger control: 6-mm circular button
c. Hybrid buttons: wide button containing a large icon
d. Plain buttons": wide button containing no icons

Experimental paradigm:

Step 1: The participant (P) completed a survey on a smartphone.
Step 2: P completed a paper questionnaire about their experience with the survey.
Guideline 19 (III)

Use 6-mm radio-button/check-box for response options

Performance measures:

a. **Question-level completion time:** time from page load to selection of “next” button
b. **Misses:** number of times a participant tapped a location on the screen that was not an active selection area

c. **Changed answers:** number of times a participant selected a different response option after their initial selection

d. **Number of categories selected:** number of categories selected for each choose-all-that-apply question

e. **Satisfaction:** rating of ease of answer selection

f. **Preference:** preference among the four designs
Guideline 19 (IV)
Use 6-mm radio-button/check-box for response options

Participants sample:

- N = 61 (Male: 18, Female 43)
- Age (year): Mean (SD) = 68.9 (4.7), Range: 59 - 75
- Education: High school or above
- Spanish origin: Yes = 5%, No = 95%
- Race: White = 74%, Black = 10%, Asian = 13%
Guideline 19 (V)
Use 6-mm radio-button/check-box for response options

Results:
*Question-level completion time*—Mixed Linear Model. No significant differences. (Figure A)
*Misses*—Significantly more misses using 2-mm circle than other designs ($p < 0.05$). (Figure B)
*Answer changes*—No significant differences. (Figure C)
*Number of categories selected*—No significant differences.
Guideline 19 (VI)

Use 6-mm radio-button/check-box for response options

Results:

**Satisfaction (ease of answer selection)** – Easy to select an answer. No significant differences. (Panel A)

**Preference** – The large circular icon was favored ($p < 0.01$). (Panel B)
Guideline 19 (VII)
Use 6-mm radio-button/check-box for response options

Evidence Summary

- Smaller button (2 mm) leads to more misses.
- Smaller button appears associated with longer completion time.
- Larger circular button (6 mm) is favored.
Guideline 15 (I)

Bold survey question stems and italicize instructions

Research questions:
1. What is the optimal combination of typographical styles for instructions and question stems?

Experimental design:
Between-subjects 2x2 factorial design: Question (Bold, Plain) and Instruction (Italic, Plain)
   a. Bold question stem and Plain instructions
   b. Bold question stem and Italicized instructions
   c. Plain question stem and Italicized instructions
   d. Plain question stems and Plain instructions
Guideline 15 (II)

Bold survey question stems and italicize instructions

<table>
<thead>
<tr>
<th>Bold + Plain</th>
<th>Bold + Italic</th>
<th>Plain + Italic</th>
<th>Plain + Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you completed a secondary (high) school diploma or equivalent?</td>
<td>Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).</td>
<td>Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).</td>
<td>Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).</td>
</tr>
<tr>
<td>Yes, secondary (high) school diploma</td>
<td>Yes, secondary (high) school diploma</td>
<td>Yes, secondary (high) school diploma</td>
<td>Yes, secondary (high) school diploma</td>
</tr>
<tr>
<td>Yes, secondary (high) school equivalency certificate</td>
<td>Yes, secondary (high) school equivalency certificate</td>
<td>Yes, secondary (high) school equivalency certificate</td>
<td>Yes, secondary (high) school equivalency certificate</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Guideline 15 (III)
Bold survey question stems and italicize instructions

Experimental paradigm:
- Completing a 5-question survey.

Performance measures:
- **Survey completion time**: Duration between the appearance of the first survey question screen and the completion of the last question in the survey.
- **Difficulty rating**: Self-report of task difficulty on a 5-point scale with 1 being very easy and 5 being very difficult.
- **Preference**: Preferred mobile survey design among the 4 designs.
Guideline 15 (V)

Bold survey question stems and italicize instructions

Participants sample:

- N = 30 (Male: 9, Female 21)
- Age (year): Mean (SD) = 69.8 (5.6), Range: 59 - 80
- Education: High school or above
- Spanish origin: Yes = 3%, No = 97%
- Race: White = 53%, Black = 23%, Asian = 24%
Guideline 15 (VI)

Bold survey question stems and italicize instructions

Results:

- **Survey completion time**: Effect of Instruction (ANOVA, $p < 0.03$) – Italics $\rightarrow$ Time ↓

- **Difficulty rating**: No difference, generally easy (multinomial logistic regression) – 1 = 86%, 2 = 10%, 3 = 2%

- **Preference**: Bold question stem = 83%, Bold question stem + Plain instructions = 52%
**Summary**

**A systematic approach:**

1. A mobile survey respondent model
2. Develop topics for standards and guidelines
3. Gathering evidences
4. Formulate standards and guidelines

**Limitations and next steps:**

- **Sample size** – At least 20 per condition for lab-based experiments, more for virtual experiments.
- **Experimental design** – Think through carefully. On point to address research questions. Between-subjects or within-subjects?
- **Simple study** – Don’t be over ambitious.
- **A living document** – Continue to incorporate new evidences, and to update the standards and guidelines accordingly.
4 standards and 23 guidelines

Contact: lin.wang@census.gov