



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)

Paul Beatty (CBSM Chief)

Benjamin Saunders (COR)

Harry Lee (former Champion)

Barbara LoPresti (former Advisor)

Alexandra Figueroa(former Sponsor)

Thomas Mathew

Sabin Lakhe

Lawrence Malakhoff

Kevin Younes

Marylisa Gareau

Rebecca Keegan

Jonathan Katz

Andrew Raim

Chin Fang Weng

Luke Larson

Joanna Fane Lineback

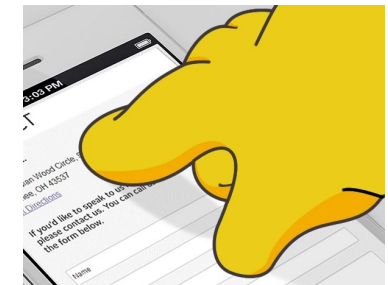
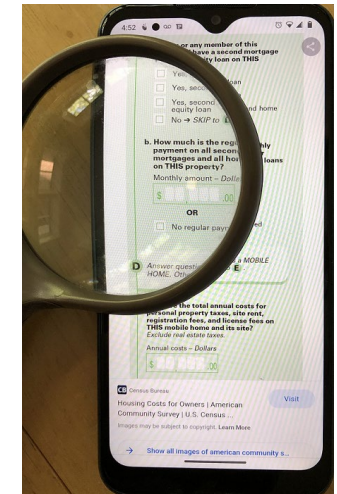
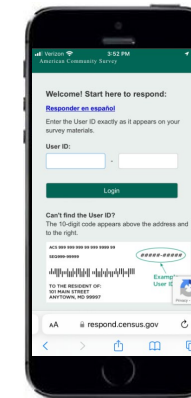
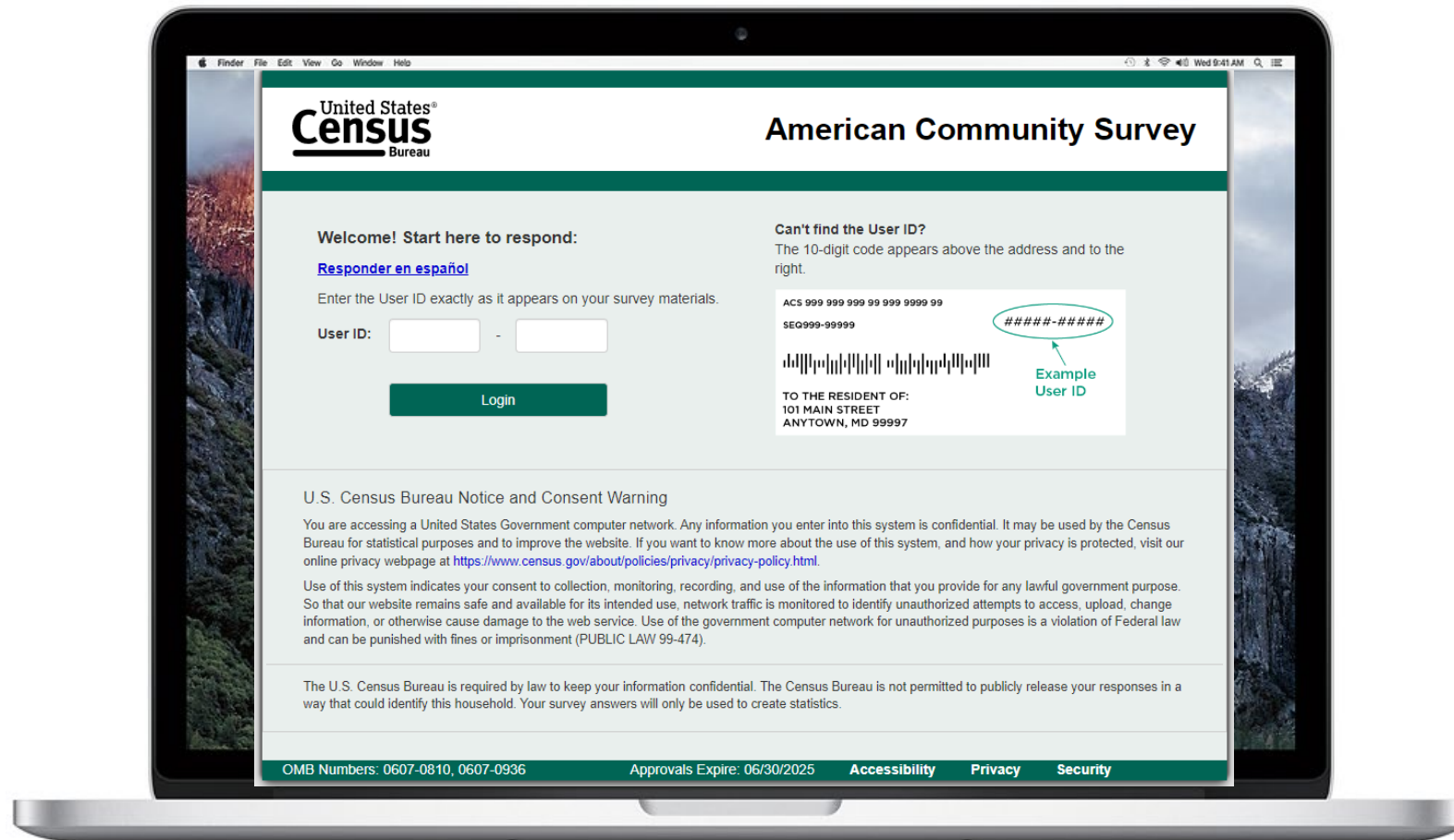
Wei Chen

Patricia Summers

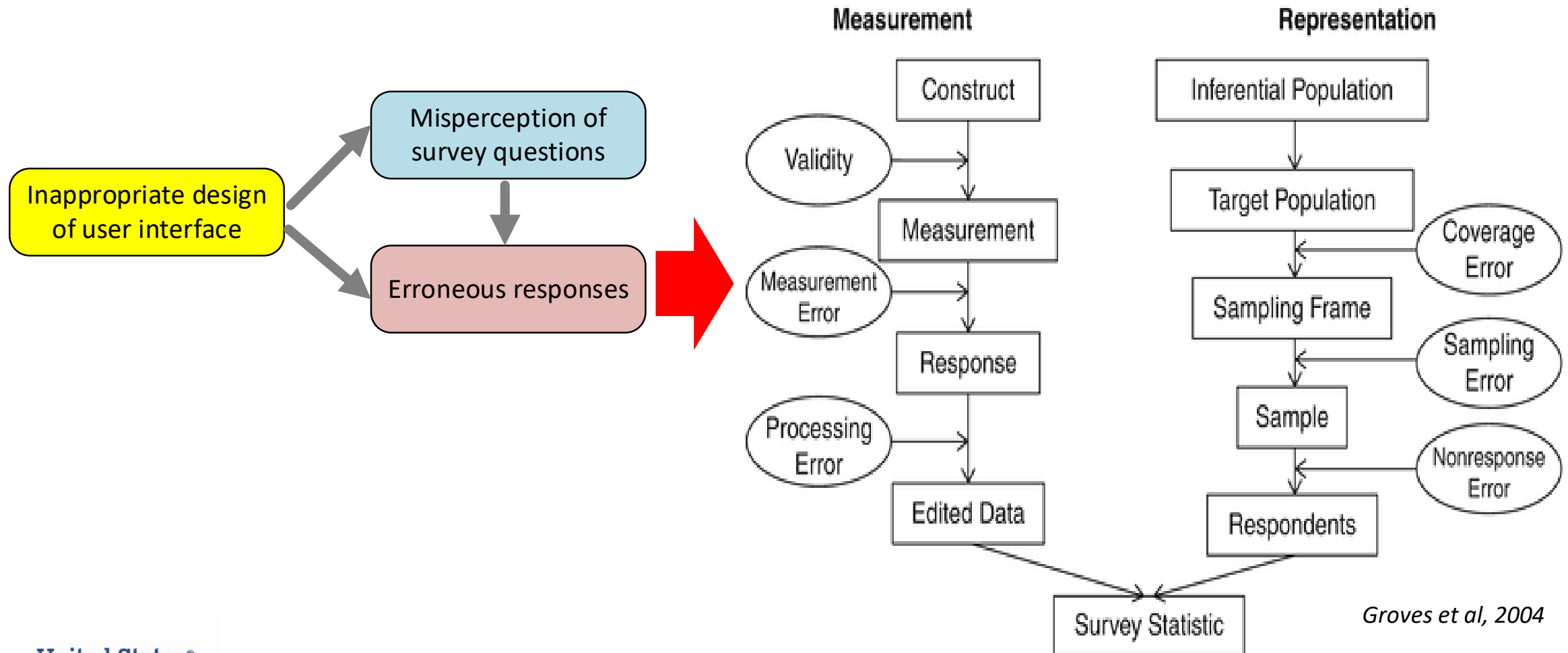
Josef Argueta

Ashley Peterson

Challenges in Mobile Survey Design



Problem



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**.

PROPOSED STANDARDS AND GUIDELINES FOR MOBILE SURVEY INSTRUMENT DESIGN

IOE 2015 BCase 01 Project Team

First Edition

March 4, 2022

Technical Lead and Editor:
Lin Wang

Contributing Authors (in alphabetical order):
Christopher Antoun, Brian Falcone, Shelley Feuer, Ivonne Figueroa,
Elizabeth Nichols, Erica Olmsted-Hawala, Alda Rivas, Lin Wang

Project Manager:
Russell Sanders

Champions:
Harry Lee, John Abowd



A Systematic Approach

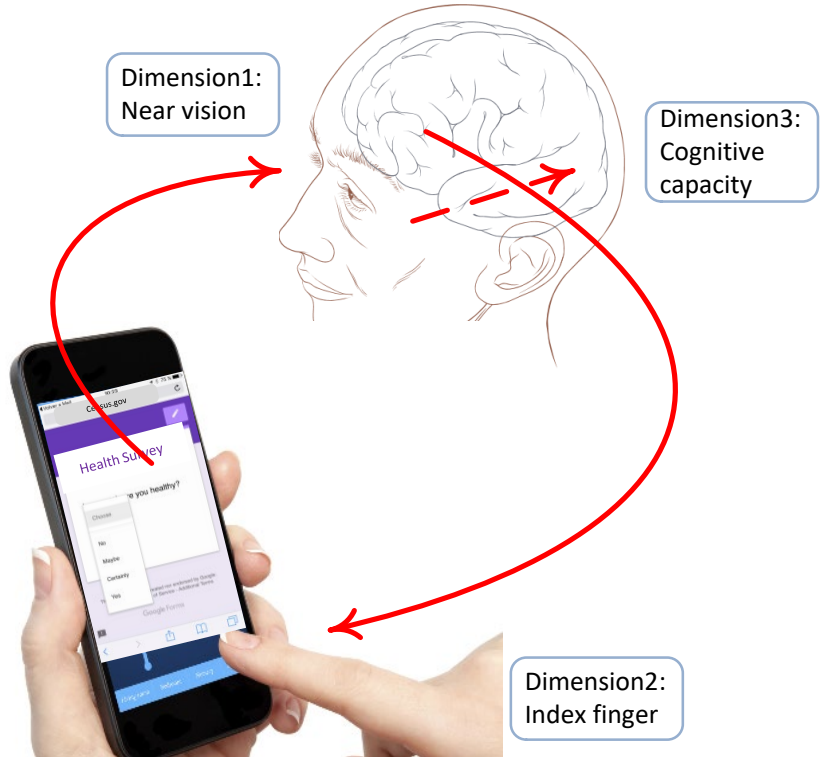
Step 1: Developing a mobile survey respondent model

Step 2: Formulating topics for standards and guidelines

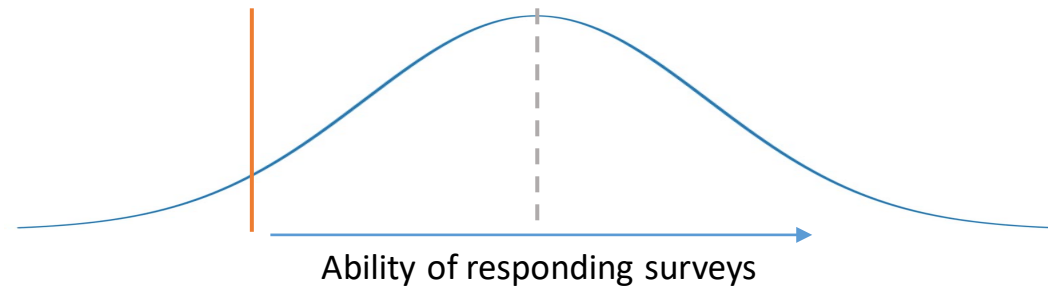
Step 3: Gathering evidences

Step 4: Making standards/guidelines based on evidence

A Mobile Survey Respondent Model



Distribution of potential respondents in terms of ability



Dimension I:
Near vision (for reading)

Habitual visual acuity: reading newspaper

Contrast sensitivity: usual (newspaper print)

Color vision: impaired

Dimension II:
Index finger

Operating fingertip breadth: 13 mm

Operating finger mobility: stiff but able to operate a smartphone

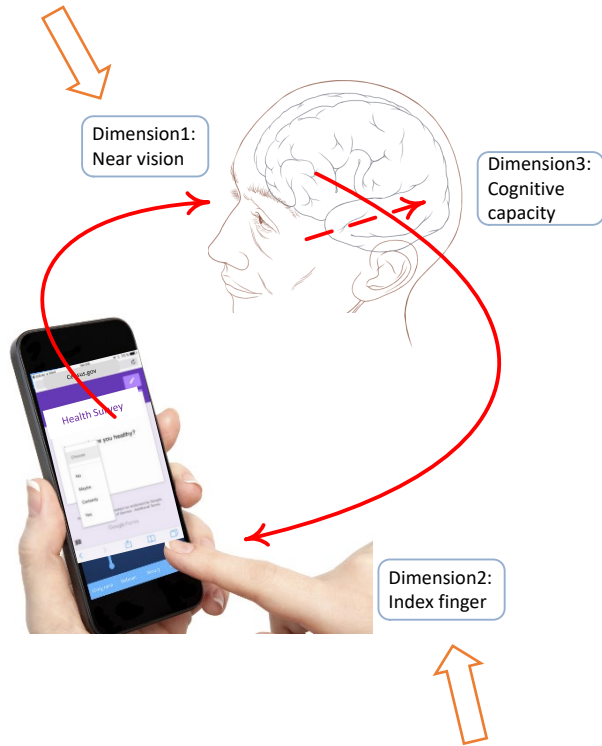
Dimension III:
Cognitive ability

Mentally alert

Language: fluent in English

Education: 8th grade or equivalent

Developing Topics for Standards



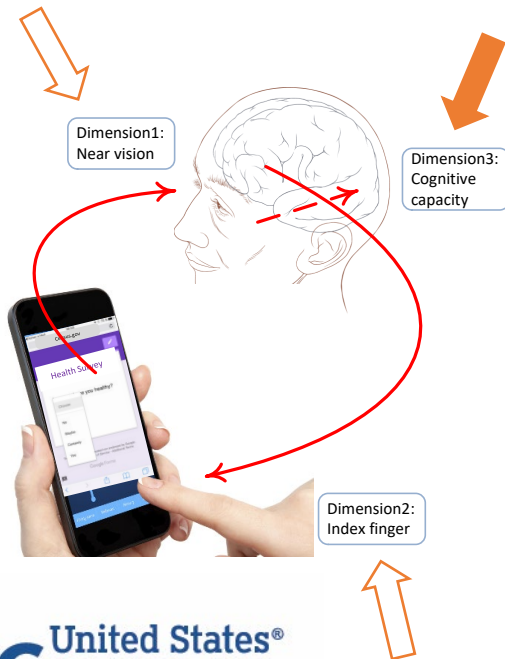
- Standards concern the basic operations across different elements of survey responses, serving as rules.
- Three areas of user interface elements:
 - 1) Touch target
 - 2) Text entry and display
 - 3) Text/graphics luminance and color

Topic candidates:

- Touch target size
- Touch target spacing
- Text field height
- Text size
- Foreground/background luminance contrast
- Color combination

Developing Topics for Guidelines

Guidelines concern specific survey response operations, serving as “best practices.”



Iterative Brainstorming

1st 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

Category	Sample Topic
Question instruction	Layout
Question stem	Text color
Response	Response options orientation
Navigation	Optimal navigation method
Support features	Within-question Help link
General	Text-Field Labeling



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.

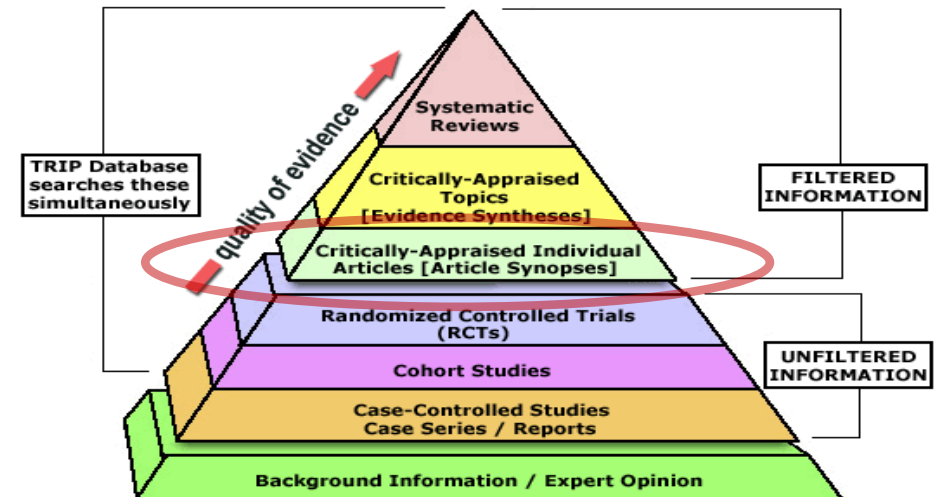
Literature Review

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

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.



EBM Pyramid and EBM Page Generator, © 2006 Trustees of Dartmouth College and Yale University. All Rights Reserved. Produced by Jan Glover, David Izzo, Karen Odato and Lei Wang.

Behavioral Experiments

Research question:

Is Design A better than B, C ...?

Performance measures:

Effectiveness, Efficiency, Satisfaction.

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

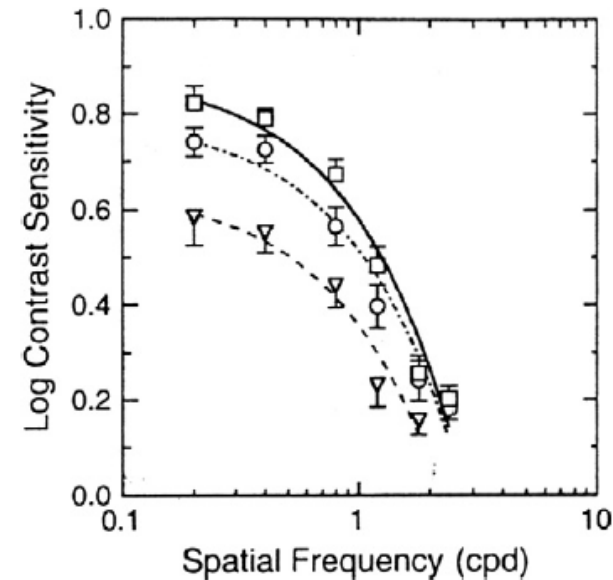
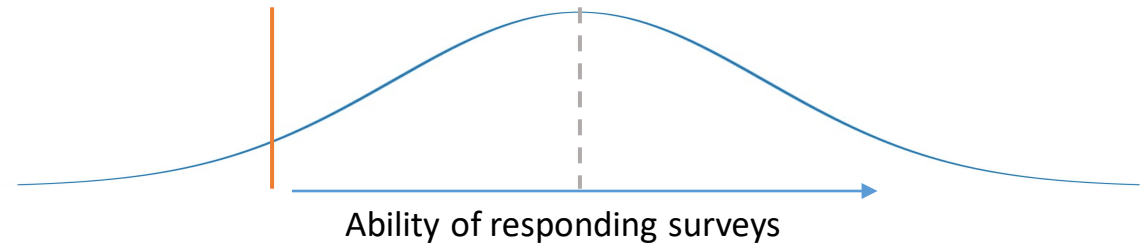
Experimental paradigm: Task design.

Experimental design: Factorial design.

Participants sample:

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

Distribution of potential respondents in terms of ability

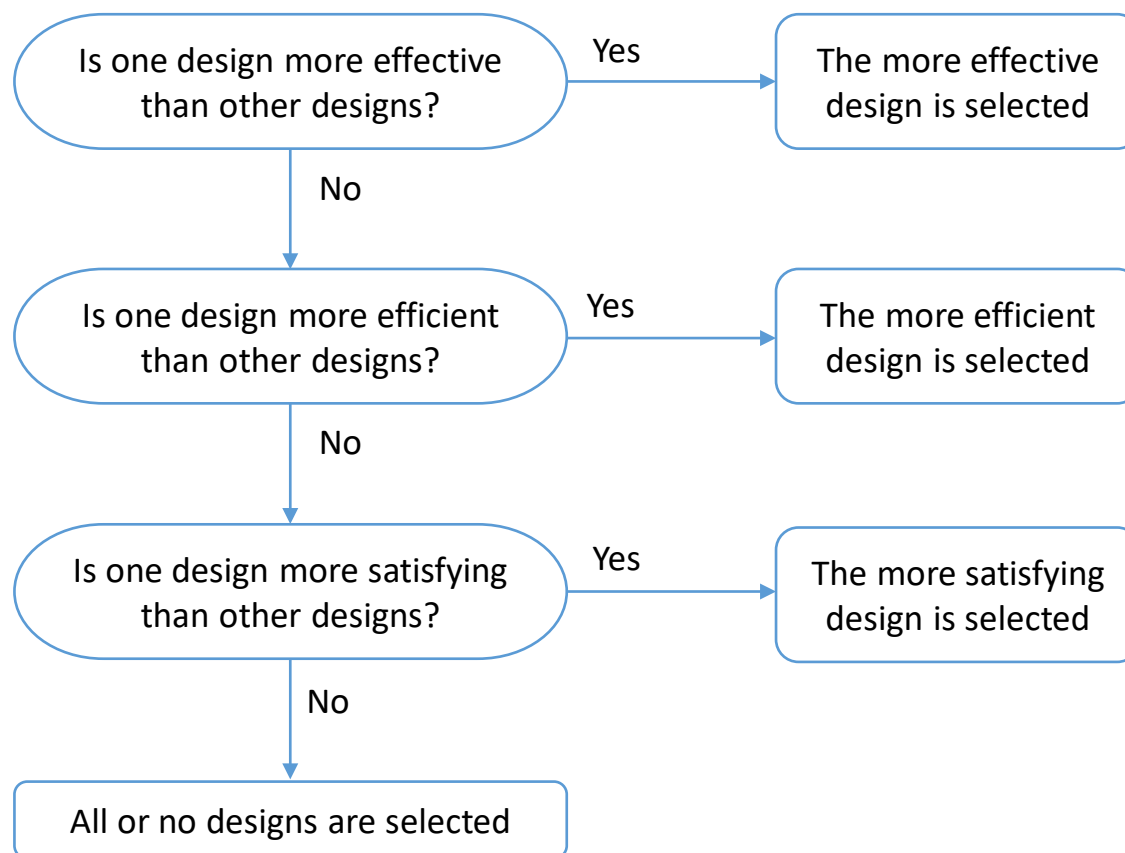


font size
font size
font size
font size

Fig. 2. From Scheffrin 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)>?





Current Version of Standards and Guidelines for Mobile Survey Design

Standards

Category	Relevant Standards
Touch target size	Standard 1
Text display	Standard 2
Luminance and color	Standard 3, 4

Guidelines

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



Sample Standards and Guidelines

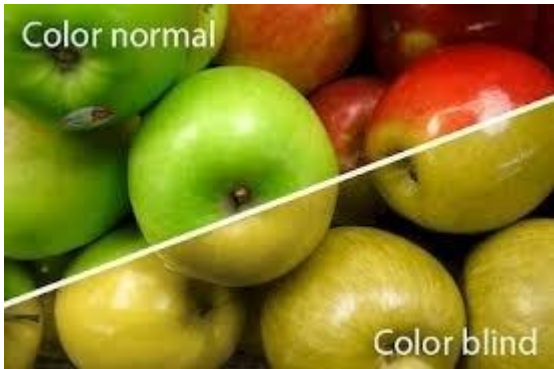
- Standard 4 – Literature review
- Standard 1 – Behavioral experiment
- Guideline 12 - Behavioral experiment
- Guideline 19 – Behavioral experiment
- Guideline 15 – Behavioral experiment



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

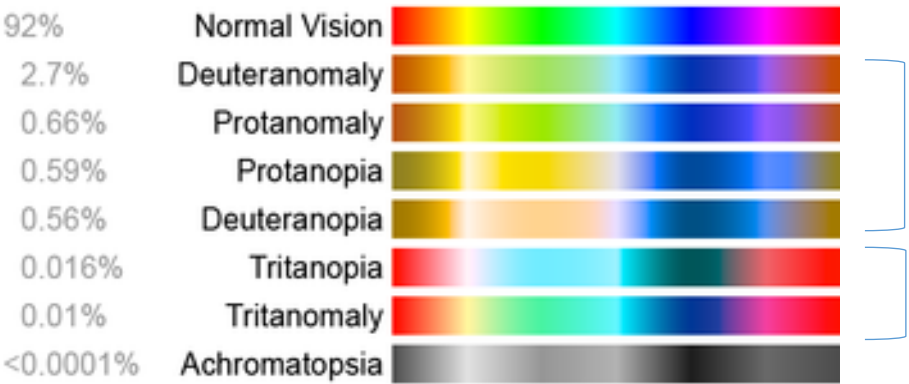


good color vision



blue-yellow deficient

Total Color Blindness



Standard 4 (II)

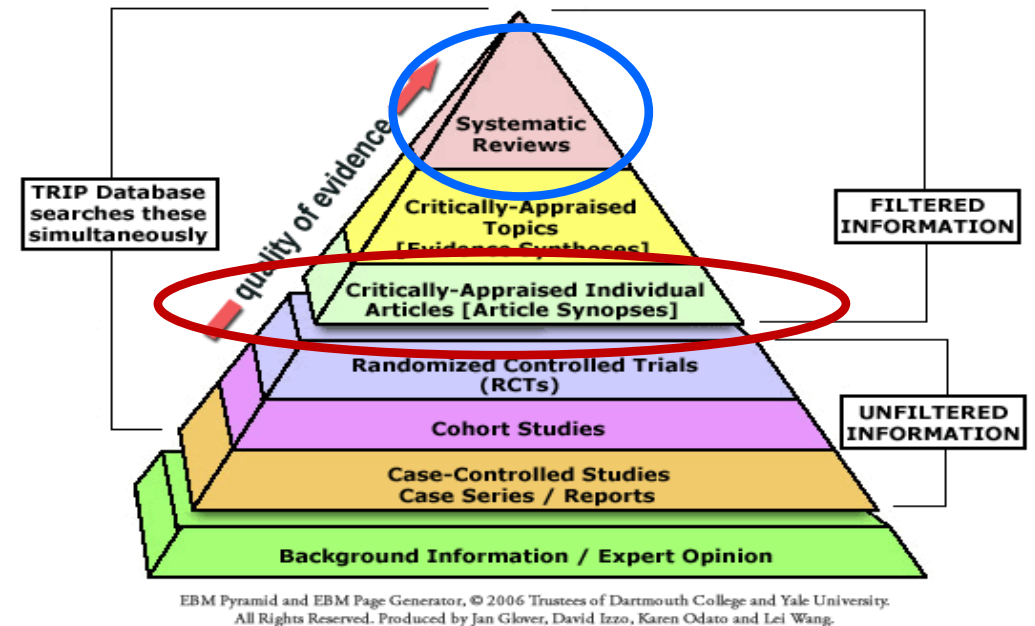
Use of Color: avoid placing red and green colors next to each other

Evidence:

A well-established clinical fact

(<https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness>)

W3C Guideline 1.4.1. In: Web Content Accessibility Guidelines (WCAG) 2.1. W3C. 2018.



Standard 1 (I)

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

Research question:

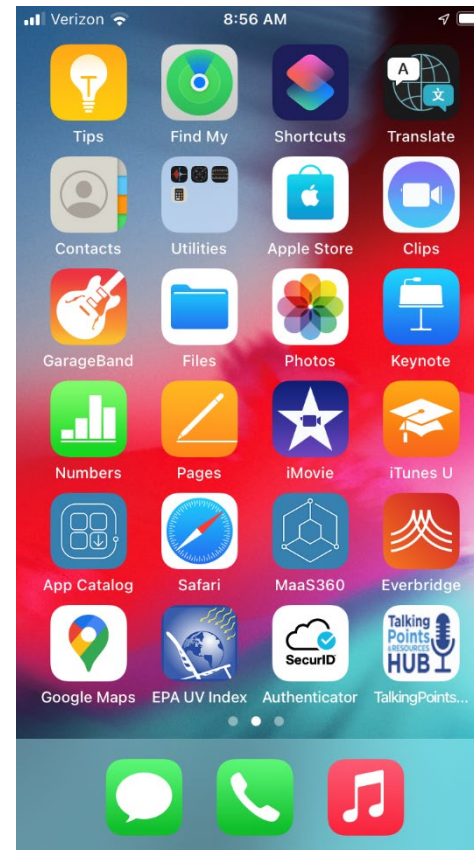
What is the optimal size of a touch target and spacing in mobile surveys?

Performance measures:

- Target touch
- Target touch time
- Task difficulty rating

Success criterion:

Target touch success rate $\geq 80\%$



We depend too much on science and not enough on faith.

- ☐ 1 = strongly disagree
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7 = strongly agree

Previous

Next

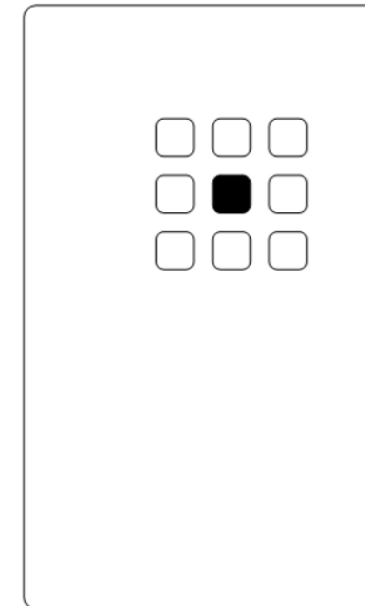
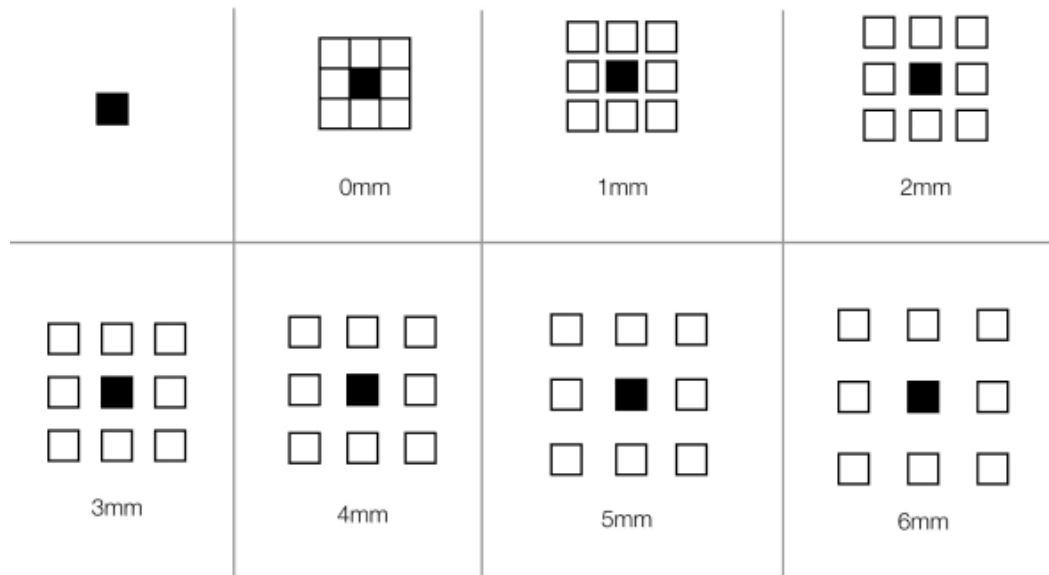
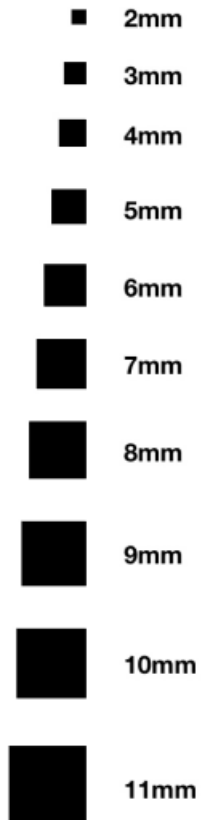


Standard 1 (II)

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.



Length of square side (mm)	Max spacing (mm)
2	8
3	7
4	7
5	6
6	6
7	5
8	5
9	4
10	4
11	3



Standard 1 (III)

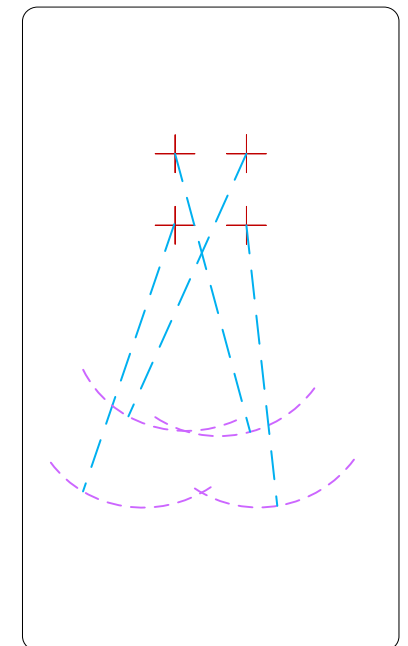
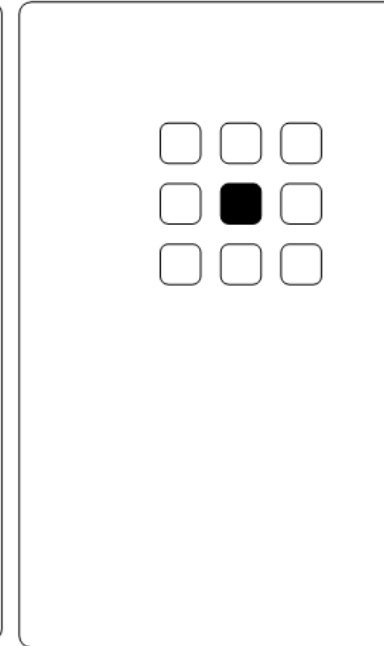
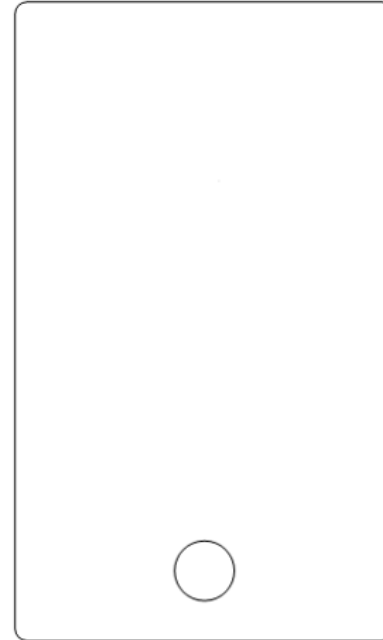
Size of Touch Button: at least 6 mm of square side or circle diameter

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

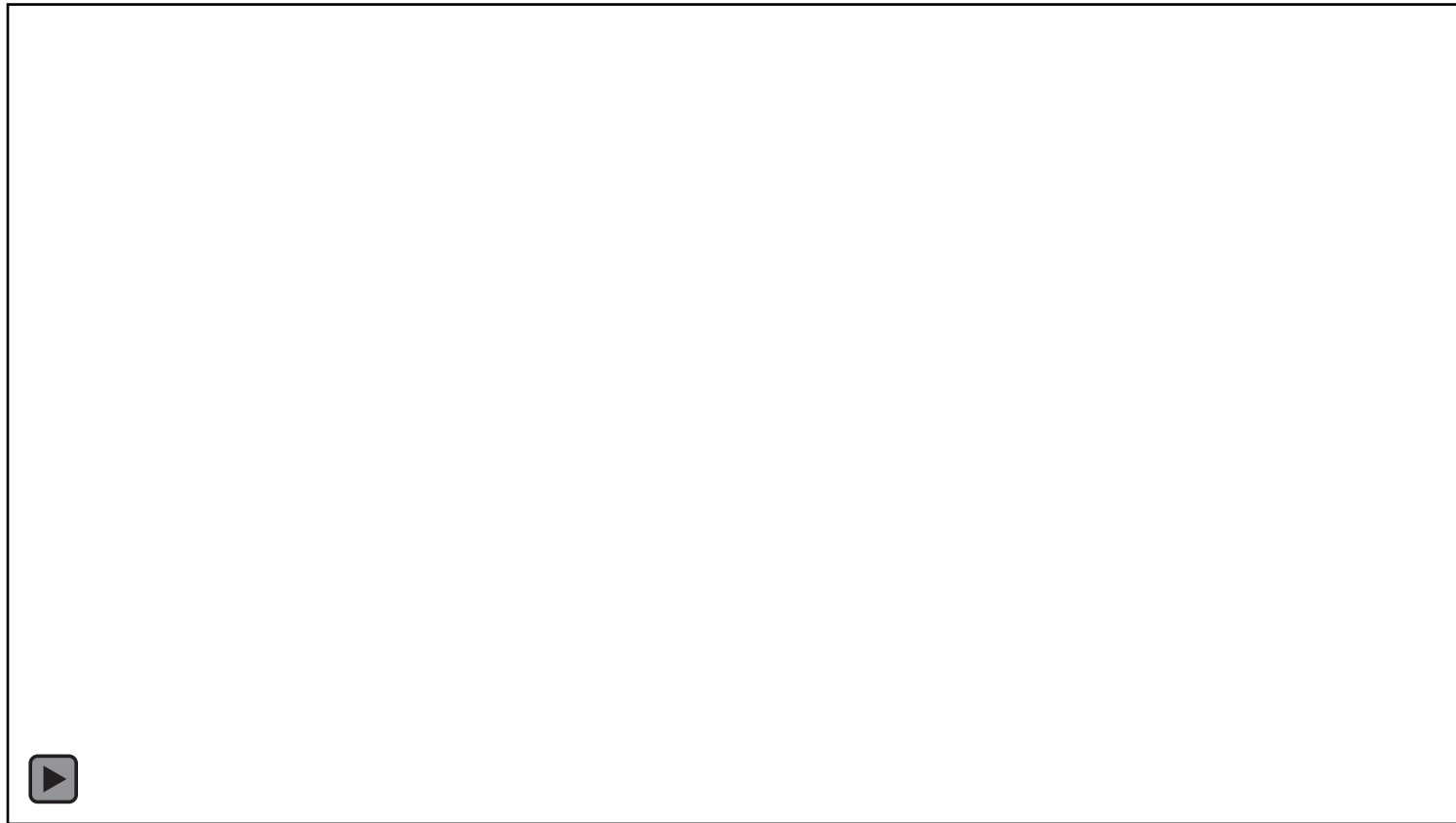
Touch Time: Time from leaving the Start button to touching the target

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?

Target Size (mm)	Target-touch Success Rate (%)
2	18.46
3	35.91
4	53.85
5	75.00
6	86.77
7	90.38
8	92.69
9	94.62
10	97.69
11	99.62



Standard 1 (VIII)

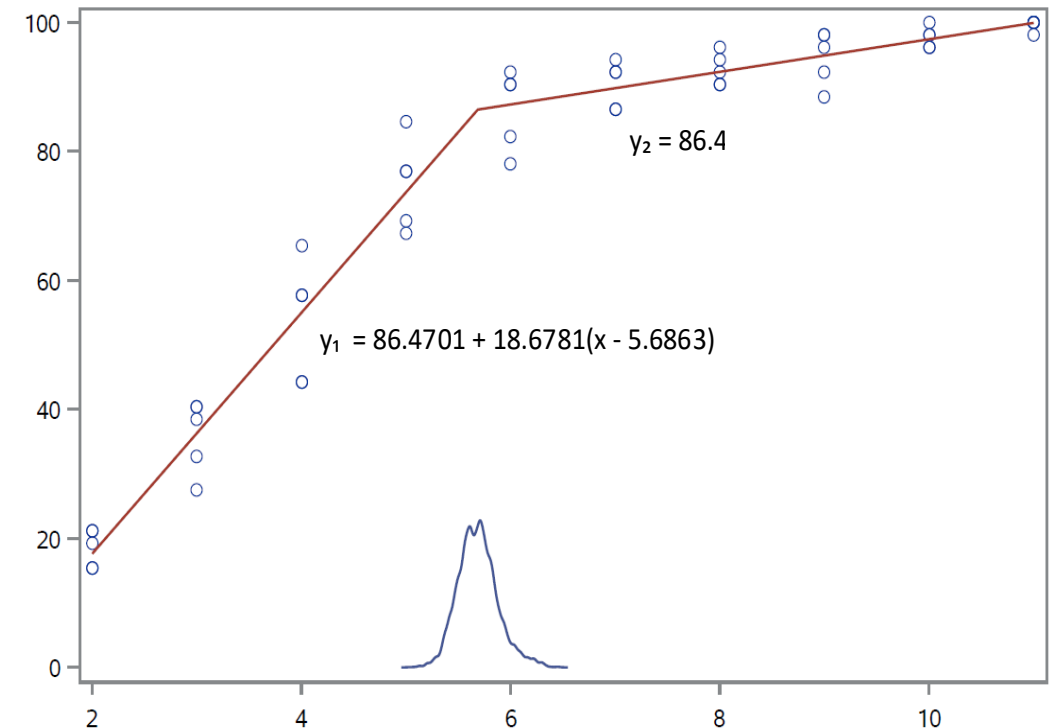
Size of Touch Button: at least 6 mm of square side or circle diameter

Change point detection

- Bayesian inference with Markov Chain Monte Carlo (MCMC) simulations.

Parameter	N	Mean	STD	95% HPD Interval	
α	20000	86.4701	2.1036	82.5055	90.6716
β_1	20000	18.6781	0.9742	16.6761	20.4971
β_2	20000	2.5285	0.5318	1.4311	3.5431
cp	20000	5.6863	0.1912	5.3359	6.0966
σ^2	20000	21.8097	4.5987	13.7218	30.9075

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

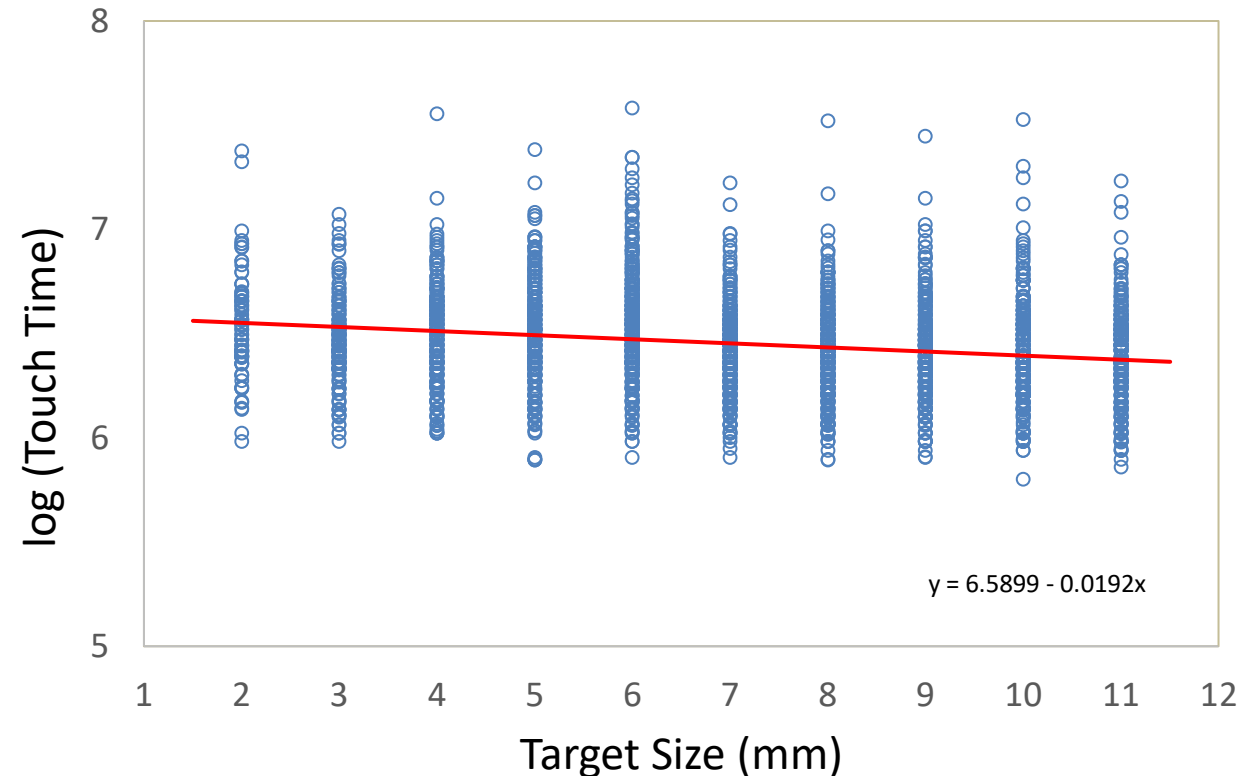


Standard 1 (IX)

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 $\uparrow \rightarrow$ touch time \downarrow ($p < 0.0001$)
- No significant effects of target spacing nor of the interaction between target size and target spacing.



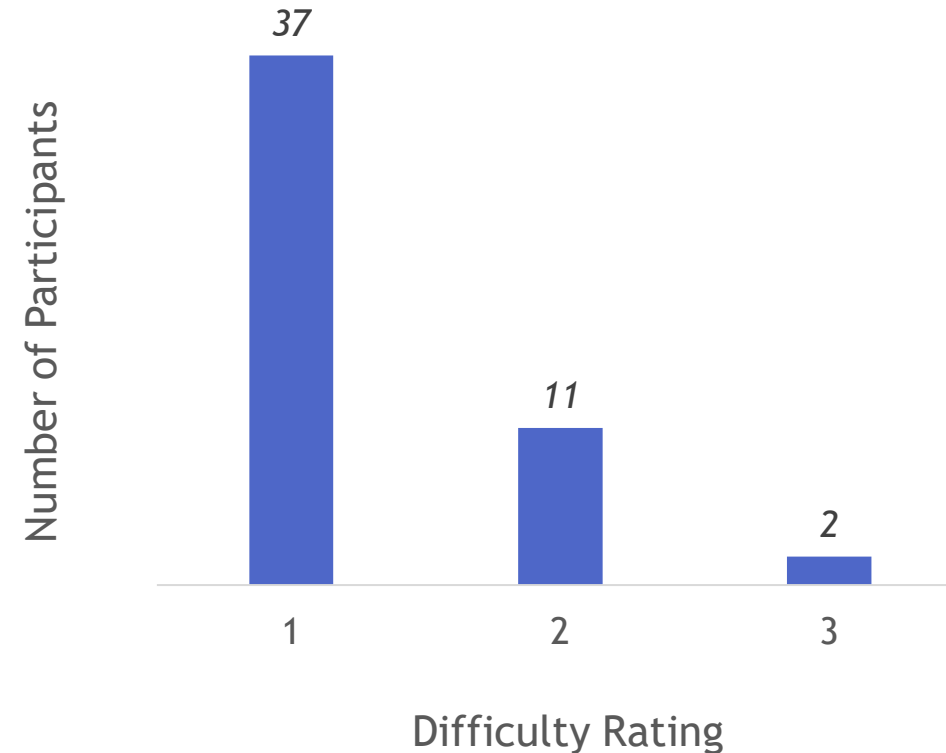


Standard 1 (XI)

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”

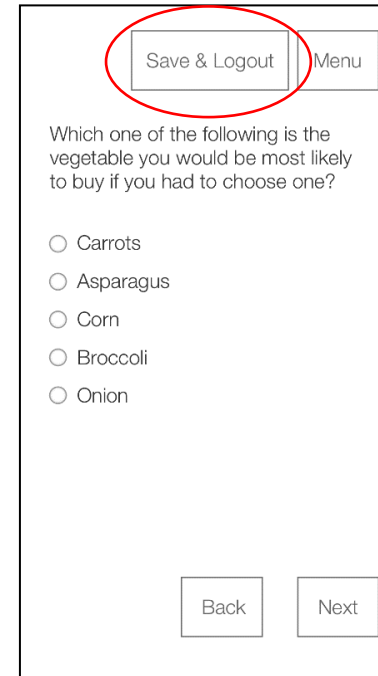
Research question:

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

Experimental design:

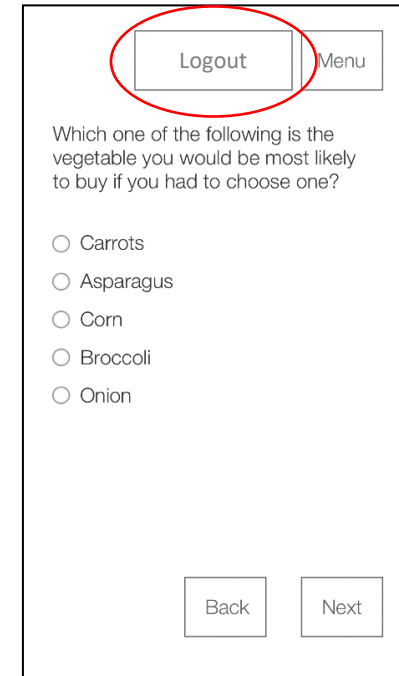
Between-subjects design with one 2-level factor, *Button Label*:

- Logout
- Save & Logout



Form A shows a survey interface with a button labeled "Save & Logout" highlighted by a red circle. The button is located at the top right of the form, next to a "Menu" button. Below the buttons is a question: "Which one of the following is the vegetable you would be most likely to buy if you had to choose one?" followed by five radio button options: Carrots, Asparagus, Corn, Broccoli, and Onion. At the bottom of the form are two buttons: "Back" and "Next".

A



Form B shows a survey interface with a button labeled "Logout" highlighted by a red circle. The button is located at the top right of the form, next to a "Menu" button. Below the buttons is a question: "Which one of the following is the vegetable you would be most likely to buy if you had to choose one?" followed by five radio button options: Carrots, Asparagus, Corn, Broccoli, and Onion. At the bottom of the form are two buttons: "Back" and "Next".

B



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”

Performance Measure Coding Schemes	
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 A

Logout button use	Condition (n = 54)	
	Save & Logout	Logout
Yes	88.9% (24)	48.1% (13)
No	11.1% (3)	51.9% (14)

Table B

Response Option	Condition (n = 54)	
	Save & Logout	Logout
Information Saved	96.3% (26)	26% (7)
Information Not Saved	3.7% (1)	74% (20)



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?

Your neighborhood:

☐ 1 = no trust at all

☐ 2

☐ 3

☐ 4

☐ 5

☐ 6

☐ 7 = complete trust

Small icon

Your neighborhood:

☐ 1 = no trust at all

☒ 2

☐ 3

☐ 4

☐ 5

☐ 6

☐ 7 = complete trust

Large icon

Your neighborhood:

☐ 1 = no trust at all

☒ 2

☐ 3

☐ 4

☐ 5

☐ 6

☐ 7 = complete trust

Wide button + icon

Your neighborhood:

☐ 1 = no trust at all

☒ 2

☐ 3

☐ 4

☐ 5

☐ 6

☐ 7 = complete trust

Wide button

Guideline 19 (II)

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

Experimental design:

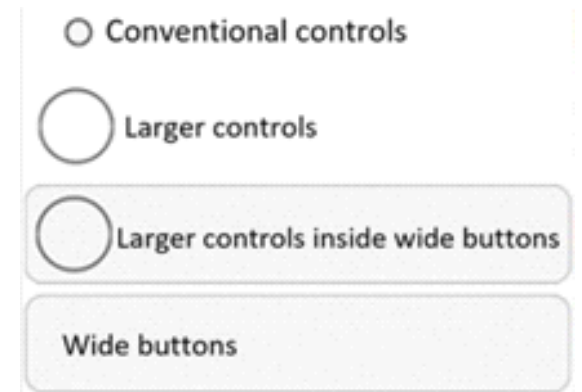
Between-subjects design with one 4-level factor, *Response Option*:

- 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

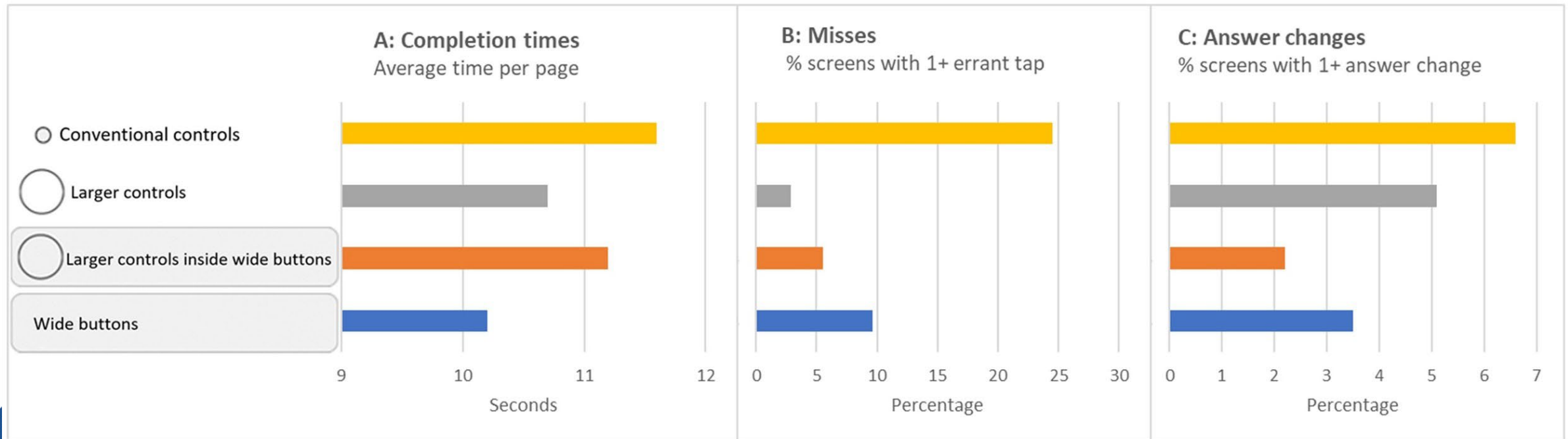
Results:

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.



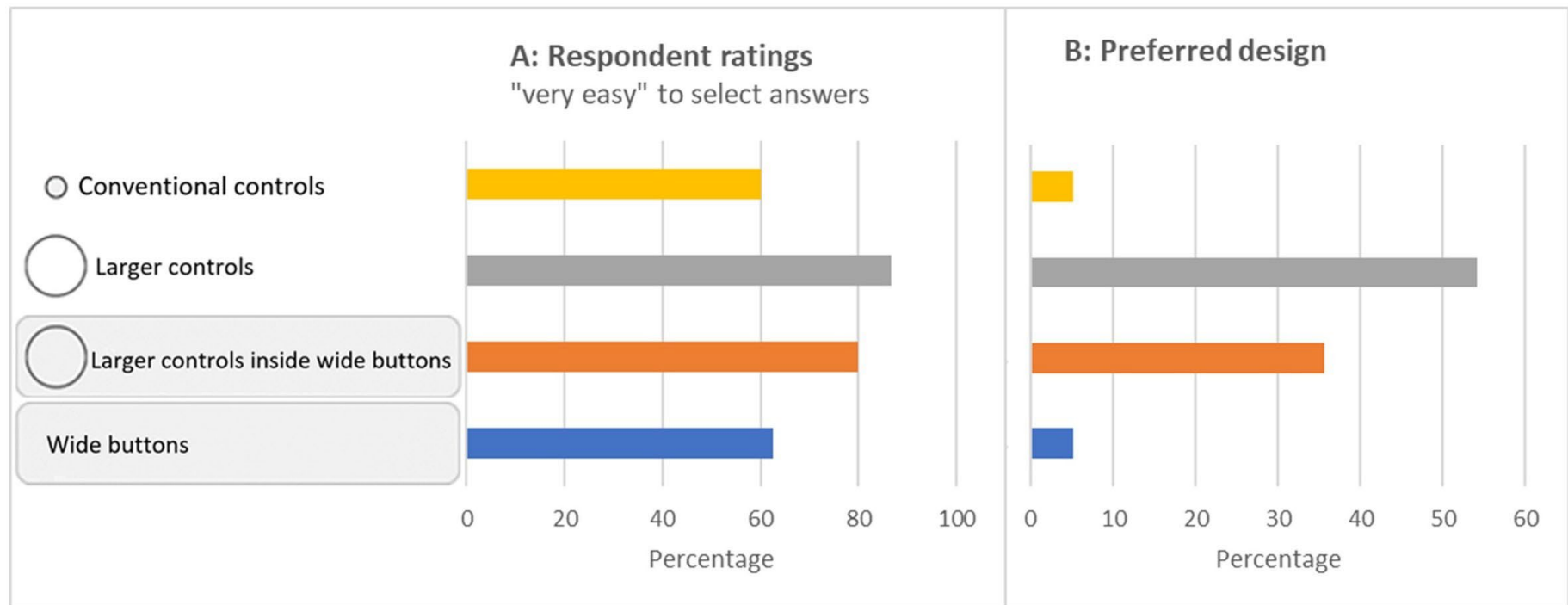
Guideline 19 (VI)

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

Bold + Plain

Have you completed a **secondary (high) school diploma or equivalent?**

Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).

☐ Yes, secondary (high) school diploma

☐ Yes, secondary (high) school equivalency certificate

☐ No

Next

Bold + *Italics*

Have you completed a **secondary (high) school diploma or equivalent?**

Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).

☐ Yes, secondary (high) school diploma

☐ Yes, secondary (high) school equivalency certificate

☐ No

Next

Plain + *Italics*

Have you completed a secondary (high) school diploma or equivalent?

Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).

☐ Yes, secondary (high) school diploma

☐ Yes, secondary (high) school equivalency certificate

☐ No

Next

Plain + Plain

Have you completed a secondary (high) school diploma or equivalent?

Examples of secondary (high) school equivalency certificates are General Educational Development (GED) and Adult Basic Education (ABE).

☐ Yes, secondary (high) school diploma

☐ Yes, secondary (high) school equivalency certificate

☐ No

Next



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%

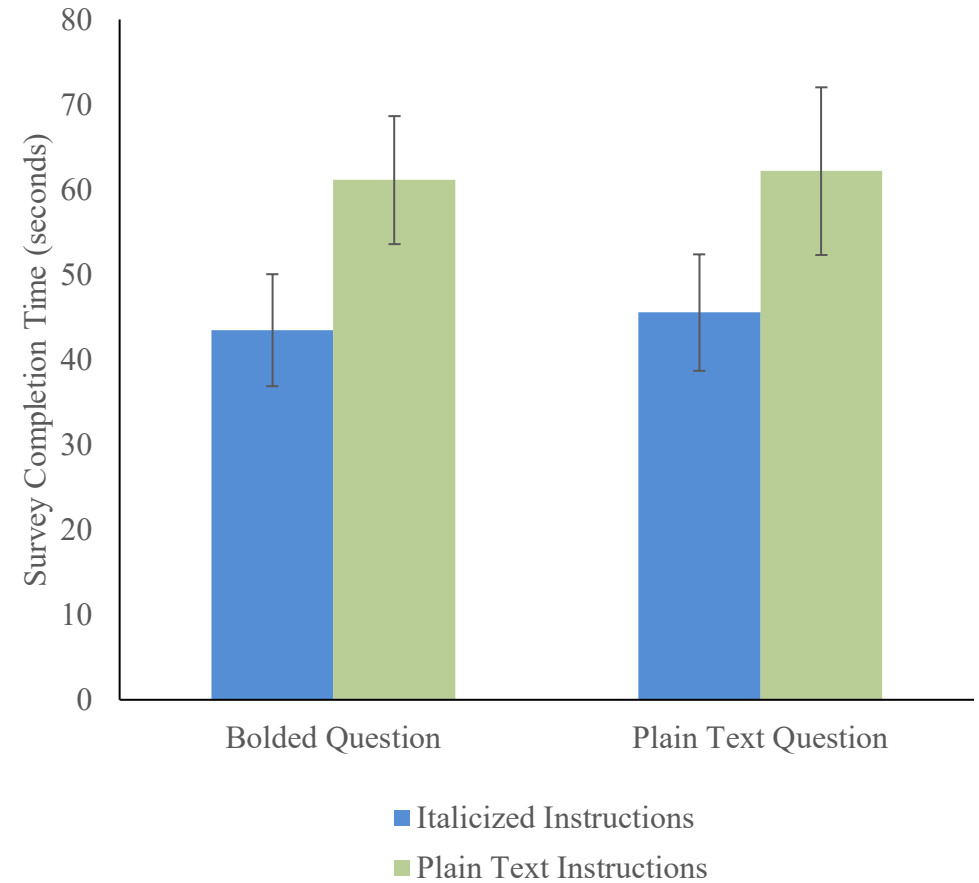


Guideline 15 (VI)

Bold survey question stems and italicize instructions

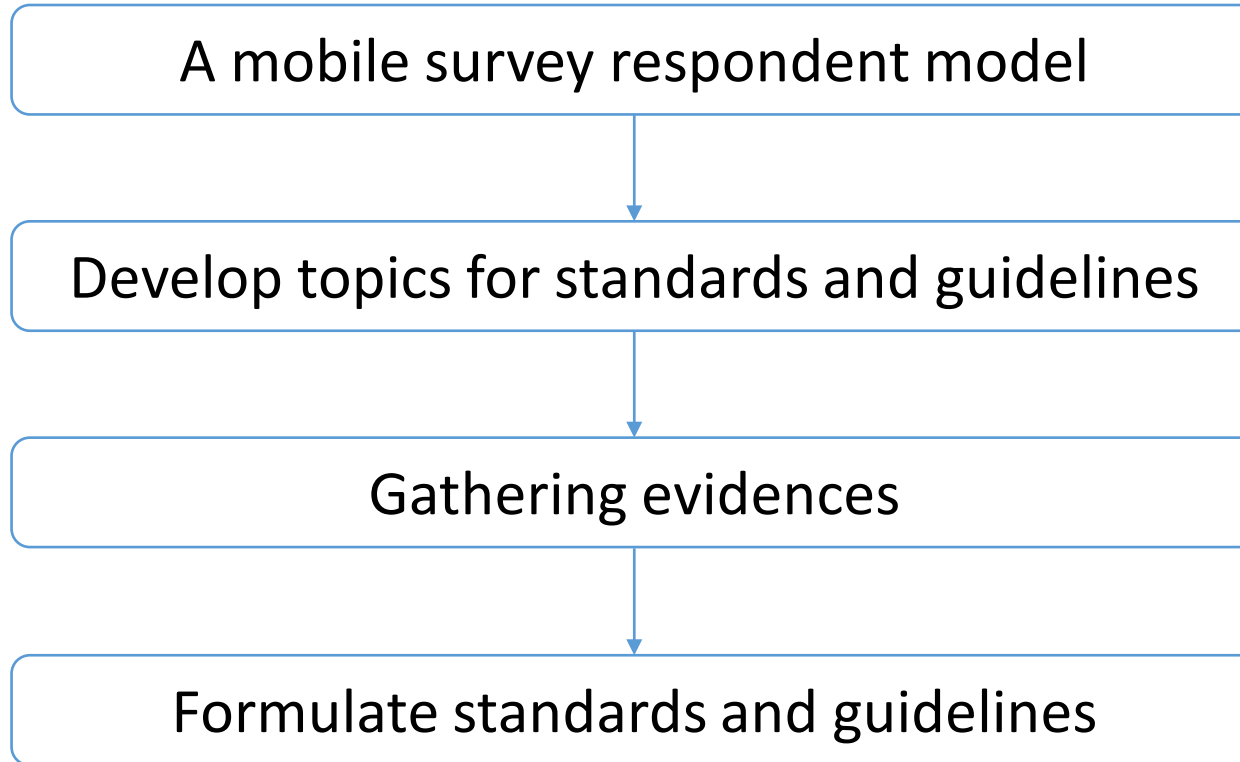
Results:

- *Survey completion time*: Effect of Instruction (ANOVA, $p < 0.03$) – **Italics** → **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:



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.

An Information Source for Survey Operations

census.gov / Library / Working Papers / 2022

Reissued 26 Apr 2022



4 standards

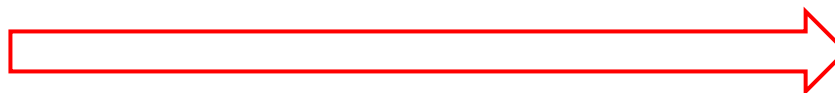


U.S. Census Bureau
Statistical Quality Standards



The leading source of quality data about the nation's people and economy

4 standards and 23 guidelines



Web Design Guidelines for
Census Bureau Surveys