Investigating Response Quality in Mobile, Tablet and Desktop Surveys:
A Comparison of Radio Buttons, Visual Analogue Scales and Slider Scales

Vera Toepoel
(Utrecht University, V.Toepoel@uu.nl)
Frederik Funke

Background

- Response quality in desktop, tablet, and mobile surveys
- Radiobuttons versus scales
  - Radiobuttons
  - Slider bars (handle, drag and drop)
  - Visual Analogue scales (point and click)
  - (combination of slider and VAS)
  » Big buttons
- Number of scalepoints (5, 7, 11)
- Evaluation of the layout
- Personal characteristics

Smartphone as research tool?

- Web surveys are completed on different devices
  - Desktop PC
  - Tablet
  - Mobile phone
- Mobile phones are different than regular desktop PCs
  - Screen size
  - Touchscreen
- 72% of the Dutch own a mobile phone with Internet access
- Only about 5% uses a smartphone for survey completion
- 57% of panel members with a smartphone used the device when prompted
  (dynamically programmed survey; Toepoel and Lugtig, published in SSCORE)

Research objectives

- Is there a difference in response quality between response formats?
- Is this related to the number of scale points?
- Is this related to the device for survey completion?
- Is the response format a significant predictor for the evaluation of the questionnaire?
- Is the above related to personal characteristics such as socio-demographics and webographics?
Method

- Data collected in the panel of GfK
  - Representative of Dutch population 15+, edu, gender
- Fieldwork in April 2014
- 3*5*3 experimental design
  - Desktop vs tablet vs mobile phone
  - Radiobuttons vs slider vs VAS vs combination slider/VAS vs big buttons
  - 5 vs 7 vs 11 point scale
    - Also added: continuous VAS
- Every format was made *identical* (e.g. length, color) except for the *functionality* (e.g. point-and-click vs drag-and-drop) and *answer type* (e.g. bar vs buttons)
- N=5,077
- Questionnaire duration: about five minutes
Response

- Fieldwork April 8-16
  - Response percentage 30%;
  - Nonresponse 34%;
  - 32% stopped because quota fulfillment;
  - 4% dropout.

<table>
<thead>
<tr>
<th></th>
<th>desktop</th>
<th>tablet</th>
<th>mobile</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitation</td>
<td>2,694</td>
<td>2,695</td>
<td>2,696</td>
<td>8,085</td>
</tr>
<tr>
<td>April 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invitation</td>
<td>2,618</td>
<td>6,228</td>
<td>8,846</td>
<td></td>
</tr>
<tr>
<td>April 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>1,709</td>
<td>1,702</td>
<td>1,666</td>
<td>5,077</td>
</tr>
</tbody>
</table>

Response

- Panelists were randomly assigned to desktop, tablet or mobile

<table>
<thead>
<tr>
<th>Assigned</th>
<th>desktop</th>
<th>tablet</th>
<th>mobile</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>desktop</td>
<td>76%</td>
<td>28%</td>
<td>19%</td>
<td>1,709</td>
</tr>
<tr>
<td>tablet</td>
<td>15%</td>
<td>54%</td>
<td>28%</td>
<td>1,702</td>
</tr>
<tr>
<td>mobile</td>
<td>9%</td>
<td>18%</td>
<td>53%</td>
<td>1,666</td>
</tr>
</tbody>
</table>
Mean scores across 16 items

- Mean scores are significantly different across formats
- Sliders show lower scores compared to radio buttons
  - Difference .18 for 5-point
  - Difference .24 for 7-point
  - Difference .48 for 11-point

- Significant differences:
  - 5-point: Tablet & Desktop
  - 7-point: Desktop
  - 11-point: Mobile

- Best design option:
  - VAS: nonsignificant from radio buttons and slider
  - Big buttons: more in line with bars than radio buttons (only significantly different in Tablet 5-point)

Item missings across 16 items

- Mean scores are significantly different across formats
- Sliders show more item missings scores compared to radio buttons
  - 5-point: .84 vs .19
  - 7-point: .61 vs .16
  - 11-point: .40 vs .16

- Significant differences:
  - 5-point: all devices
  - 7-point: all devices
  - 11-point: Mobile

- Best design option:
  - Radio buttons or Big buttons: lowest item missings
  - VAS: low item missings and nonsignificant from other formats

Not Applicable

- Only significant in 7-point scale
  - Again, more in slider

- Not related to devices

- Best design option:
  - Again, better no slider

Evaluation of the questionnaire (1)

- Sum score 5 evaluation questions
- Regression analysis

- Model 1: Questionnaire characteristics and interactions:
  - R Square = .031

- More positive:
  - 11-point

- Less positive:
  - Mobile (although optimally designed)
  - People who self-selected into device

- Interactions positive:
  - Big buttons * 11-point
  - Slider * Mobile
  - VAS * Mobile
  - Slider/VAS * Mobile
  - Selfselect * Mobile

- No negative Interactions
Evaluation of the questionnaire (2)

- Sum score 5 evaluation questions
- Regression analysis
- Model 2: Questionnaire characteristics and socio-demographics:
  - R Square = .067
- More positive:
  - Slider, VAS, Slider/VAS
  - 11-point
- Less positive:
  - Tablet and Mobile

- All socio-demographics significant:
  - Pos: Age
  - Neg: gender, oldest old, high edu, high income, single

- Very few sig. Interactions
  - Male prefer big buttons
  - High edu dislike 11-point

Evaluation of the questionnaire (3)

- Sumscore 5 evaluation questions
- Regression analysis
- Model 3: Questionnaire characteristics and webographics:
  - R Square = .033
- More positive:
  - Slider, Slider/VAS
  - 11-point
- Less positive:
  - Mobile

- Webographics:
  - Pos: Survey experience desktop and mobile
  - Neg: general experience with mobile
  - Not significant:
    - general experience with desktop and tablet,
    - survey experience with tablet,
    - completion time

- No significant interactions!

Conclusions

- Do not use Sliders!
- Use VAS!
  - Respondents evaluate questionnaire better
  - No significant differences in mean scores compared to buttons and slider
  - Item nonresponse is (almost) as low as buttons
- Respondents evaluate 11-point better
- Respondents evaluate bars better
- Respondents evaluate mobile worse
  - Mobile survey experience seems to make respondents milder
- Almost no significant interactions with personal characteristics and webographics

Thank you

- Questions: mail me!
  - v.toepoe@uu.nl