Analyzing Cognitive Burden of Survey Questions with Paradata: A Web Survey Experiment

JAN KAREM HÖHNE¹
STEPHAN SCHLOSSER¹
DAGMAR KREBS²

General Online Research Conference
02 - 04 March 2016, Dresden (Germany)
Research Question

- Do Agree/Disagree (AD) questions require higher cognitive burden than Item-Specific (IS) questions?
  - So far, empirical evidence is missing.

What are possible **REASONS** for differences?
Survey Practice

- Since Likert (1932) AD questions have become popular.
  - German General Social Survey, Eurobarometer, ISSP …
  - European Sociological Review, Political Analysis, Social Science Quarterly …

- Reasons for popularity¹:
  - Measuring different contents with the same scale.
  - Streamlining questionnaires.

¹ Saris et al. (2010)
Critique on Survey Practice

- Fowler (1995) identifies several drawbacks of AD questions:
  - Insufficiently anchored at one end of the content continuum.
  - Complex cognitive processing.
  - Small discriminatory power of response categories.
  - Prone to response bias.

⇒ It seems to be more appropriate to use IS questions.
**Research Hypotheses**

**Hypothesis 1:**
AD questions produce longer response times than IS questions.

**Hypothesis 2:**
AD questions produce a higher amount of answer changes than IS questions.

**Hypothesis 3:**
AD questions show a higher degree of low response quality than IS questions.
Both groups received eight questions dealing with “work” and “competition” and 5-point end labeled response scales.¹

**Syllables equivalence:**
Questions including response categories do not differ in more than 2 syllables.²

¹ One question per screen.
² see Lenzner et al. (2010)
Sample

- The web survey was conducted at two German universities in May 2015.

<table>
<thead>
<tr>
<th>Sample Size:</th>
<th>N = 1005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex:</td>
<td>54% female</td>
</tr>
<tr>
<td>Age:</td>
<td>Mean = 24.9 (SD = 4.2)</td>
</tr>
<tr>
<td>Education:</td>
<td>College preparatory secondary school or college degree</td>
</tr>
<tr>
<td>Survey Experience:</td>
<td>93% have participated in a web survey once before</td>
</tr>
</tbody>
</table>

*Chi-square tests reveal no significant differences between the groups with respect to age, gender, and survey experience.*
Paradata System

- JavaScript based system that is implemented in Unipark.
- Paradata are captured by an invisible user-defined question.
- Submitting a “page” of the web survey, paradata are stored with survey data.
- The system records ...
  - response times, mouse clicks/movements, scrolling, screen size in pixels, leaving web survey browser ...
  - Reaching a pixel-perfect, cross-browser scripting with CSS and HTML.
Outlier Definition Procedure

1st Step
Participants who left survey were defined as outliers.

2nd Step
“Common” outlier procedure for remaining participants.¹

\[
\begin{align*}
Q_{0.5} + (3 \times (Q_{0.75} - Q_{0.5})) \\
Q_{0.5} - (3 \times (Q_{0.5} - Q_{0.25}))
\end{align*}
\]

¹ Hoagland et al. (2000)
Results

Hypothesis 1: AD questions produce longer response times than IS questions.

Notes: *p < .001. The t-test was calculated for the logarithmic response times. Additionally, we calculated Cohen’s d. The significance level is based on the results of the t-test. N = 919.
Results

Hypothesis 2: AD questions produce a higher amount of answer changes than IS questions.

Notes: We calculated Cohen’s $d$ as measure of effect size. $N = 919$. 
**Results**

*Hypothesis 3:* AD questions show a lower response quality than IS questions.

<table>
<thead>
<tr>
<th>Agree/Disagree Item-Specific</th>
<th>Speeding</th>
<th>Dropouts</th>
<th>Item Non-Responses</th>
<th>Primacy Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.8% (61)</td>
<td>2.4% (28)</td>
<td>.3% (3)</td>
<td>13.1% (120)</td>
</tr>
<tr>
<td></td>
<td>2.8% (26)</td>
<td>2.7% (31)</td>
<td>.5% (5)</td>
<td>13.2% (121)</td>
</tr>
</tbody>
</table>

\( \chi^2 = 11.45; df = 1^* \)

\( \chi^2 = .77; df = 1 \)

\( \chi^2 = .76; df = 1 \)

\( \chi^2 = 1.25; df = 1 \)

*Notes:* *p < .001. For speeding we used the lower 10\textsuperscript{th} percentile of all response times. For primacy effects we used the number of responses given on the first half of the response scale. \( N = 919. \)
Limitations

- Study is based on a student sample.
  - Generalizability of results.

- (Highly) motivated participants.
  - Response rate (4.9%).\(^1\)

\(^1\) see Couper et al. (2004)
Conclusion

- Processing of IS questions seems to be more complex – response times.
- No differences regarding mapping difficulties – answer changes.
- Small differences with respect to response quality.
- AD questions seem to be more prone to a superficial processing.

⇒ Altogether, results disprove established assumptions about cognitive effort associated with AD and IS question.
Thank you for your attention!

Contact: jhoehne@uni-goettingen.de
**Literature**


