Identifying and Dealing with Item Nonresponse in Open-ended Questions in a Cross-national Context

Katharina Meitinger
(GESIS - Leibniz Institute for the Social Sciences)
Lars Kaczmirek
(University of Vienna, Australian National University)

CSDI, Limerick. 28 March, 2018
Overview

- Open-ended questions and the issue of nonresponse
- Detecting and reducing nonresponse in open-ended questions
- Results:
  - Initial nonresponse
  - Nonresponse after conversion
- Conclusion
Open-ended Questions

- Important source of information to:
  - Ask about exploratory topics
  - Receive answers when the list of possible response categories is not known or is too long (Fowler 1995)
  - Assess validity and comparability with probing (e.g., Schuman 1966; Behr et al. 2017)

- Open-ended questions are cognitively more demanding for respondents than closed items and increase response burden (Bradburn 1978)
  - Risk of increased item-nonresponse at open-ended questions
How to reduce nonresponse?

Prevent occurrence of nonresponse by optimizing visual design:
- Size of answer boxes (e.g., Christian and Dillman 2004; Smyth et al. 2009; Behr et al. 2014)
- Examples (Tourangeau et al. 2014)
- Use of interactive elements (e.g., Emde and Fuchs 2012)

Convert nonrespondents to provide substantive responses:
- Forced answer → Risk of break-off
- Repetition of open-ended question
- Motivational sentences (see also Oudejans & Christian 2011; Zuell et al. 2014)
We can distinguish between different types of nonresponse that differ regarding the respondents’ strength of intention to avoid answering a question:

- **One word only**
  - Example: economy

- **Don’t knows**
  - Example: I have no idea, I can’t make up my mind

- **Other nonresponse**
  - Example: my personal experience, it depends

- **Too fast response** (Response took less than 2 seconds)

- **Refusals**
  - Example: no comment

- **Complete nonresponse**

- **No useful answer**
  - Example: dfgjh frtrt
How to detect & convert nonresponse?

**Precondition of nonresponse conversion:** Detect instances of NR:
- Complete nonresponse: Easy to detect (empty answer box)
- Remaining nonresponse types: Harder to detect, needs coding

**EvalAnswer:**
- Tool for automatic detection of different types of nonresponse
- Conversion attempts: Repetition of open-ended question & motivational sentence
- Languages: German, English, Spanish
- Free tool available (https://git.gesis.org/surveymethods/evalanswer)
- **Kaczmarek, Meitinger, & Behr (2017):** Working paper on technical implementation

*Higher data quality in web probing with EvalAnswer: a tool for identifying and reducing nonresponse in openended questions*
Kaczmarek, Lars; Meitinger, Katharina; Behr, Dorothée
EvalAnswer: Automatic Nonresponse Detection and Conversion

How much do you agree or disagree with the following statement?

I feel more like a citizen of the world than of any country.

- agree strongly
- agree
- neither agree nor disagree
- disagree
- disagree strongly
- can't choose

Please explain why you selected "neither agree nor disagree".

The statement was: "I feel more like a citizen of the world than of any country."

Automatic nonresponse detection

We would like to understand what you had in mind when you answered the original question. Please try to answer this follow-up question:

Please explain why you selected "neither agree nor disagree".

The statement was: "I feel more like a citizen of the world than of any country."

Closed question

Open-ended probe

NR conversion attempt: Repetition of probe & motivational sentence
RESEARCH QUESTIONS

RQ1: How prevalent is nonresponse in our data?

RQ2: How prevalent are the different NR types and are there cross-national differences?

RQ3: Can we convert nonrespondents?

RQ4: Do countries and NR types differ regarding their conversion level?
Data

- Two web surveys with panelists from non-random online access panels (Study 1: N=2,685; Study 2: N=2,689)
- Country sample: Germany, Great Britain, the U.S., Spain, and Mexico
- Quotas for age (18-30, 31-50, 51-65), gender, and education (lower education vs. higher education)
- Data collection in May and June 2014
- Replication of questions from International Social Survey Programme Modules (2012, 2013, and 2014)
- 29 open-ended questions (probes) with 35,252 responses
- Automatic nonresponse-detection and conversion (EvalAnswer)
RQ 1
How prevalent is nonresponse in the five countries?
Nonresponse

- Overall nonresponse (29 questions): 9.45%

Mean Percent NR By Country

- NR less prevalent in Mexico and Spain than in other countries
RQ 2
How prevalent are the different NR types and are there cross-national differences?
Overall Nonresponse Distribution by NR Types

- "1 word," "don’t know," "not useful" most frequent NR types
Country differences regarding prevalence of NR types:
- Germany: “not useful”
- GB: “don’t know”
- U.S.: “refusal”
- Mexico: “empty answer box”
RQ 3

Can we convert nonrespondents?
Conversion rates - Overall

We can approximately reduce nonresponse by half
RQ 4
Do countries and NR types differ regarding their conversion level?
Conversion rates – By Country

Nonresponse conversion by country

<table>
<thead>
<tr>
<th>Country</th>
<th>NR before convers</th>
<th>NR after convers</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>US</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>GB</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td>SP</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>MX</td>
<td>0.03</td>
<td>0.01</td>
</tr>
</tbody>
</table>

→ Nonresponse conversion works in all countries
Conversion rates: By NR Type
% of Initial Nonresponse

→ There are differences in conversion rates along NR Types
→ “Unwilling” NR types more difficult to convert
“1 word” & “other” most successfully converted in all countries
“not useful”: lowest conversion rates
“empty answer box”: Germany lowest, Mexico highest conversion
“Don’t know”: conversion works best in Mexico
Nonresponse...

- differed in prevalence across countries (Mexico and Spain lowest NR)
- differed in prevalence across NR types:
  - “1 word,” “don’t know,” “not useful” most frequent NR types
  - “empty answer box” account for only 11% of all NR types → likely there is an underestimation of NR in classical methodological studies
- differed in prevalence across NR types and countries:
  - Germany: “not useful”; GB: “don’t know”; U.S: “refusal”; Mexico: “empty answer box”

- NR conversion:
  - reduces NR by half and works in all countries
  - NR Types: “Unwilling” respondents more difficult to convert
  - Similarities in conversion across countries: easiest converted were “1 word” & “other” successfully; most difficult to convert: “not useful”
  - Differences in conversion:
    - “empty answer box”: Germany lowest, Mexico highest conversion
    - “Don’t know”: Mexico highest conversion
Next steps & things to consider

- Develop more targeted approaches to reduce nonresponse that address specific nonresponse types
- Necessary to avoid pushing respondents too far
- Understanding the cultural specificities of nonresponse
- Extending tool by adding additional languages
Thank you!

Contact: katharina.meitinger@gesis.org
Success of nonresponse conversion attempts by variable
Logistic regression

Number of obs = 35,252
Wald chi2(7) = 501.37
Prob > chi2 = 0.0000
Pseudo R2 = 0.0608

Log pseudolikelihood = -10360.978

(Std. Err. adjusted for 5,374 clusters in case_id)

| nonresponse1 | Coef.   | Robust Std. Err. | z       | P>|z|   | [95% Conf. Interval] |
|--------------|---------|------------------|---------|-------|----------------------|
| age          | -0.0245711 | 0.0020299        | -12.10  | 0.000 | -0.0285497 to -0.0205925 |
| men          | 0.1188542  | 0.0608898        | 1.95    | 0.051 | -0.0004876 to 0.238196  |
| edu_high     | -0.4528201 | 0.0618897        | -7.32   | 0.000 | -0.5741217 to -0.3315185 |
| DE           | 1.67613    | 1.125621         | 14.89   | 0.000 | 1.455512 to 1.896748   |
| SP           | 0.6440151  | 0.1230056        | 5.24    | 0.000 | 0.4029284 to 0.8851017  |
| GB           | 1.444733   | 1.137312         | 12.70   | 0.000 | 1.221824 to 1.667642   |
| US           | 1.552755   | 1.153124         | 13.47   | 0.000 | 1.326747 to 1.778763   |
| _cons        | -2.360749  | 1.293158         | -18.26  | 0.000 | -2.614204 to -2.107295 |