Mix and Match: Question Structure and Self-Proxy Agreement Across Household Type

Angie O’Brien, Dave Tuttle, and Jonathan Katz, U.S. Census Bureau

Disclaimer: This presentation is intended to inform people about research and to encourage discussion. The views expressed are those of the authors and not those of the U.S. Census Bureau. The presentation has been reviewed for disclosure avoidance and approved under clearance number: CBDRB-FY22-CBSM002-031
Literature

• Does a single informant know enough to be able to accurately report on the characteristics of the entire sampled unit?
• Literature = mixed
  • (Briggs 1992; Fulton et al., 2020; Tourangeau & Yan; Zuckerbraun, et al. 2020)
• How much effort should survey researchers expend to obtain data from each person from a sampled unit? (Moore, 1990)
Public Participation in the Arts (PPA)

• Supplement of the Current Population Survey (CPS)
• Purpose = to measure adult participation in various arts and cultural activities
  • CATI/CAPI
  • Approx. 56,000 households (2017)
  • Approx. every five years
• Allows for self and within household proxy responses
  • 2017 → 2 respondents
  • 2022 → 1 respondent
Exploratory Research Questions

1. Is there evidence that certain question structures are more conducive to reliable proxy reporting than others?
   • Yes/No
   • Scale
   • Open-Ended Numeric
   • Select-All

2. Does the type of relationship between the proxy and the target respondent affect proxy reliability?
   • Related
   • Unrelated
Methods

• 18 pairs of 36 participants
  • Household members
    • Related (16)
    • Unrelated (10)
• Winter/spring 2021
• National sample
• Proxy questions
  • E.g., During the last 24 months, did _____ visit an art museum or gallery?
    • → Target respondent

Sample Demographic Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>34</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>26</td>
</tr>
<tr>
<td>Black or African American</td>
<td>10</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Completed High School</td>
<td>3</td>
</tr>
<tr>
<td>Some College, no degree</td>
<td>5</td>
</tr>
<tr>
<td>Associates Degree</td>
<td>4</td>
</tr>
<tr>
<td>Bachelor's Degree</td>
<td>16</td>
</tr>
<tr>
<td>Post-Bachelor's Degree</td>
<td>8</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>7</td>
</tr>
<tr>
<td>25-34</td>
<td>8</td>
</tr>
<tr>
<td>35-44</td>
<td>10</td>
</tr>
<tr>
<td>45-54</td>
<td>3</td>
</tr>
<tr>
<td>55-64</td>
<td>2</td>
</tr>
<tr>
<td>65+</td>
<td>6</td>
</tr>
<tr>
<td>Relationship</td>
<td></td>
</tr>
<tr>
<td>Related</td>
<td>26</td>
</tr>
<tr>
<td>Unrelated</td>
<td>10</td>
</tr>
</tbody>
</table>
Proxy Match Rates by Question Structure
Exact Match

• When both the proxy and the target respondent provide the same answer to a question (Fulton, et al. 2020)

• How many times did you do this ____ visit an art museum or gallery during the last 24 months?
  • Proxy: 2 times
  • Target: 2 times
Near Match

• When proxy and target respondent do not choose the same response, but are only one response category apart from each other (Fulton, et al., 2020)

• How many times did you do this ____ visit an art museum or gallery during the last 24 months?

  • Proxy: 2 times
  • Target: 3 times
# Question Structure Details

<table>
<thead>
<tr>
<th>Question structure</th>
<th>Number of questions</th>
<th>Average # of response options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes/no</td>
<td>90</td>
<td>2</td>
</tr>
<tr>
<td>Scale</td>
<td>10</td>
<td>4.4</td>
</tr>
<tr>
<td>Open-ended numeric</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>Select-all</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>
Exact Proxy Match Rates by Question Structure

![Bar chart showing proxy match rates by question structure.]

- Yes/No: 70%
- Scale: 50%
- Open-Ended Numeric: 10%
- Select-All: 0%
Near Proxy Match Rates by Question Structure

- Scale: 90%
- Open-Ended Numeric: 10%
- Select-All: 50%
Relationship Type
Relationship type

- Related
  - Spouses
  - Unmarried partners
  - Parent/child
  - Siblings
- Unrelated
  - Housemates
- Social distance
  - (Bickart et al., 1990; Katz et al., 2022)
Yes/No Average Exact Match Rate by Relationship Type

- Related: 68.9% (n = 26)
- Unrelated: 67.6% (n = 10)
Yes/No Average Exact Match Rate by Relationship Type

- Parent/Child
- Siblings
- Unmarried Partners
- Spouses
- Unrelated Housemates
Conclusions

• In proxy reporting, the number of response options matters
• Proxy/target respondent agreement may be more nuanced than exact match rates suggest
• Relationship type = not the best predictor of proxy reliability
  • Social distance matters
Future Research

• Statistical design

• Additional testing on related vs unrelated proxies
  • E.g., spouses vs. parent/child

• Different ways to do complex questions
  • E.g., select-all → yes/no

• Aggregate
  • Compare target and proxy data quality in the aggregate (e.g., item missingness and variance, etc.)
Thank you 😊