THE EFFECT OF QUESTION CHARACTERISTICS ON RESPONDENT AND INTERVIEWER BEHAVIORS
ALLYSON L. HOLBROOK, TIMOTHY P JOHNSON, YOUNG IK CHO, SHARON SHAVITT, NOEL CHAVEZ & SAUL WEINER
INTERVIEWER ADMINISTERED SURVEYS

- A form of conversation (e.g., Schwarz, 1996) and a social interaction (e.g., De Santis, 1980)
  - What is said. (e.g., Groves & McGonagle 2001 fallback statements)
  - Paralinguistic aspects of communication (tone, pitch, pace; e.g., Conrad et al., 2010)
  - What is done (nonverbal; e.g., Hornik, 1988 touching)
- Interviewers can influence both respondents’ decision to participate and responses and behavior during the survey interview
- Respondents can influence interviewers’ behaviors (e.g., via requests or errors)
- Both interviewers and respondents may be affected by question characteristics
- Both interviewer and respondent behaviors may change over the course of an interview.
  - Fatigue
  - Learning (primarily respondents)
- Difficult to assess because there are a lot of moving parts to a dyad interaction, particularly in this case where aspects of the communication are dictated by a third party (via the questionnaire and interviewer instructions)
RESEARCH QUESTIONS/HYPOTHESES

- How do question characteristics, interviewer behaviors and respondent behaviors relate?
  - H1a: Question characteristics will affect comprehension difficulties, such that respondents will have more comprehension difficulties in response to difficult questions.
  - H1b: Question characteristics will affect mapping difficulties, such that respondents will have more mapping difficulties in response to difficult questions.
  - H1c: Question characteristics will affect interviewer reading difficulties.
  - H1d: Interviewers’ reading difficulties will at least partly explain the effect of question difficulty on respondent comprehension and mapping problems.

- H2: Both question sensitivity and respondent laughter will predict interviewer laughter.
- H3: Interviewer and respondent laughter will be associated with more respondent problems.
How do interviewer and respondent behaviors change over the course of an interview?

- H4a: Respondent fatigue: Respondents will show greater problems later in an interview than they do earlier.
- H4b: Respondent learning: Respondents will show fewer problems later in an interview than they do earlier.
- H5: Interviewer fatigue: Interviewers will show greater problems later in an interview than they do earlier.

- H6a: Interviewer reading times will decrease as the interview progresses.
- H6b: Response latencies will decrease as the interview progresses.
- H6c: Faster interviewer reading speed will be associated with faster respondent answers.
- H6d: Faster interviewer reading speed will be associated with more respondent difficulties.
606 Respondents recruited via telephone
- White: N=151
- Black: N=151
- Mexican American: (N=151; English/Spanish)
- Korean American (N=150; English/Korean)

Came into the lab for an interview
- PAPI Instruments
- CAPI interview about health
- Biophysical measures
- Up to 296 health-related questions
- 77 minutes on average

Separated into five sections
- Section I: Physical Health
- Section II: Physical and Mental Health Knowledge
- Section III: Mental Health
- Section IV: Proxy Selection and Proxy Reporting for Selected Household Member
- Section V: Demographics

Order of sections 1-3 was rotated: allowed us to assess effect of position independent of specific questions

Data “stacked” so that question is the unit of analysis

More than 170,000 units for analysis (not considering missing data for individual variables)
MEASURES OF INTERVIEWER AND RESPONDENT BEHAVIOR:

- Interviewers were video and audio recorded
- Audio recordings were used to code the verbal behaviors of both respondents and interviewers
  - Up to three interviewer and three respondent behaviors were coded for each question
  - 20% of all behavior codes were validated (n=55,253 unique codes); validation rate was 95.7%
- Reading time was assessed for each question (excluding demographics)
- Response latencies were also assessed (excluding demographics)
RESPONDENT BEHAVIOR CODING VARIABLES:

- Comprehension problems (10,737 times)
  - 1 = any comprehension problem
  - 0 = no comprehension problem

- Mapping problems (14,454 times)
  - 1 = any mapping problem
  - 0 = no mapping problem

- Respondent laughs (6,018 times)
  - 1 = R laughs
  - 0 = R did not laugh

- Any problem (34,176 times)
  - 1 = any kind of behavior that indicated a problem (comprehension, mapping, memory)
  - 0 = no problem of any kind
INTERVIEWER BEHAVIOR CODING VARIABLES:

- Reading problems (28,668 times)
  - 1 = any reading problem with that question
  - 0 = no reading problem with that question
- Interviewer laughs (3,568 times)
  - 1 = I laughs
  - 0 = I did not laugh
- Any interviewer problem (32,110)
  - 1 = any kind of behavior that indicated a problem interviewer behavior
  - 0 = no interviewer problem of any kind
PROCEDURES FOR MEASURING QUESTION READING TIME AND RESPONSE LATENCIES

- The amount of time it took a respondent to answer a question after the interviewer had finished asking it
- Procedure modified from Bassilli (1996b) using three screens for each question.
  - ‘Q screen’ (question screen) - Interviewers read the question and pressed ‘Enter’ immediately after finishing. (measure of question reading time)
  - ‘R screen’ (response screen) – Interviewer entered response. (measure of response latency)
  - ‘L screen’ (response latency screen) - interviewer recorded whether the response latency was valid
- Only response latencies that were “valid” were analyzed. (82%)
- Mean question reading time was 10.95 seconds
- Mean response latency was 4.11 seconds
- Both question reading times and response latencies were transformed by taking the square root to adjust for non-normal distribution.
QUESTION CHARACTERISTICS VARIABLES: (MANIPULATED TO BE CROSSED)

Question length: Number of sentences
Text difficulty: Fleisch grade level

Question type
*Open-ended numeric response
Agree/disagree
Yes/no
Substantive categorical
Unipolar Scale (fully labeled)
Fully Labeled Single Item Bipolar Scale – with midpoint
Fully Labeled Single Item Bipolar Scale – without midpoint
Semantic differential – endpoints only labeled
Semantic differential – endpoints and midpoint labeled
Initial branching item – with midpoint
Initial branching item – without midpoint
Faces scale

Type of Judgment
Subjective
Personal characteristics or behavior
*Objective factual knowledge

Explicit DK or no opinion option?
0 = no
1 = yes

Preceded by a DK filter question?
0 = no
1 = yes

DK filter question?
0 = no
1 = yes

Time qualified?
0 = no
1 = yes

Was a show card used?
0 = no
1 = yes

Designed to be difficult?
0 = No
1 = Yes

Proxy report?
0 = No
1 = Yes

Sensitive question or known to have SD connotations?
0 = not at all
1 = somewhat sensitive or SD connotations
2 = strong sensitive or SD connotations
Average of four coders

Level of abstraction
0 = concrete
1 = moderately abstract
2 = very abstract
Average of four coders' ratings
ANALYSIS:

- Cross classified multilevel modeling approach
  - Behavior codes responses and response and question latencies are nested within question and respondent simultaneously.
  - This approach corrects for nonindependence at both question and respondent levels.
  - Can be used to analyze respondent and question characteristics as well as interviewer and respondent behaviors at the question level.
  - HLM was used to conduct analysis.
  - All analyses control for all question characteristics reported in methods and respondent demographics.
  - Questions examining changes over the course of the interview are limited to those from sections which were rotated.
## RESULTS: QUESTION CHARACTERISTICS, INTERVIEWER READING PROBLEMS, AND RESPONDENT COMPREHENSION AND MAPPING PROBLEMS

<table>
<thead>
<tr>
<th>Predictors</th>
<th>R Comprehension Problems</th>
<th>R Mapping Problems</th>
<th>I Reading Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td></td>
<td></td>
<td><strong>.78</strong></td>
</tr>
<tr>
<td>Fleish Grade Level</td>
<td></td>
<td></td>
<td><strong>.11</strong></td>
</tr>
<tr>
<td>Question abstraction</td>
<td></td>
<td>-.34**</td>
<td>-.34**</td>
</tr>
<tr>
<td>Time qualified</td>
<td>.19**</td>
<td>.19**</td>
<td></td>
</tr>
<tr>
<td>Designed to be difficult</td>
<td>2.44**</td>
<td>2.44**</td>
<td></td>
</tr>
<tr>
<td>Includes dk option</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preceded by a dk filter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is a dk filter question</td>
<td>2.01**</td>
<td>2.01**</td>
<td>1.09+</td>
</tr>
<tr>
<td>Showcard</td>
<td>.96**</td>
<td>.95**</td>
<td>.67**</td>
</tr>
<tr>
<td>Question type (numeric open ended)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree-disagree</td>
<td>-.78**</td>
<td>-.78**</td>
<td>1.22**</td>
</tr>
<tr>
<td>Yes-no</td>
<td>-1.06**</td>
<td>-1.06**</td>
<td>-.59**</td>
</tr>
<tr>
<td>Categorical</td>
<td>-1.23**</td>
<td>-1.26**</td>
<td>-.83**</td>
</tr>
<tr>
<td>Unipolar scale</td>
<td>-1.61**</td>
<td>-1.62**</td>
<td>-1.03**</td>
</tr>
<tr>
<td>Bipolar scale – no midpoint</td>
<td>-1.27**</td>
<td>-1.27**</td>
<td>-.91**</td>
</tr>
<tr>
<td>Bipolar scale – midpoint</td>
<td>-1.82**</td>
<td>-1.82**</td>
<td></td>
</tr>
<tr>
<td>Sem. Diff. – endpoint only</td>
<td></td>
<td>-.135**</td>
<td>-.135**</td>
</tr>
<tr>
<td>Sem. Diff. – endpoint &amp; midpoint</td>
<td>-2.70**</td>
<td>-2.70**</td>
<td>-1.93**</td>
</tr>
<tr>
<td>Faces scale</td>
<td></td>
<td>-1.52**</td>
<td>-1.51**</td>
</tr>
<tr>
<td>Question sensitivity</td>
<td></td>
<td>.42**</td>
<td>.42**</td>
</tr>
<tr>
<td>Proxy report</td>
<td>-.89**</td>
<td>-.89**</td>
<td>.63**</td>
</tr>
<tr>
<td>Question Position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer reading problems</td>
<td>- - - -</td>
<td>-.001*</td>
<td>-.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IMPLICATIONS

- Support for H1a, H1b, and H1c
- No support for H1d

- Question characteristics affected respondent comprehension and mapping problems and interviewer reading errors
- Interviewer reading errors, however, did not mediate the effect of question characteristics on respondent problems,
**H2: Both question sensitivity and respondent laughter will predict interviewer laughter. - YES

**H3: Interviewer and respondent laughter will be associated with more respondent and problems. - YES

<table>
<thead>
<tr>
<th></th>
<th>Interviewer Laughter</th>
<th>All Respondent Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question sensitivity</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td>Respondent laughter</td>
<td>4.42**</td>
<td>1.14**</td>
</tr>
<tr>
<td>Interviewer laughter</td>
<td></td>
<td>1.02**</td>
</tr>
</tbody>
</table>

**p<.01
HOW DO INTERVIEWER AND RESPONDENT BEHAVIORS CHANGE OVER THE COURSE OF AN INTERVIEW?

<table>
<thead>
<tr>
<th>Question Position</th>
<th>Respondent Comprehension Problems</th>
<th>Respondent Mapping Problems</th>
<th>All Respondent Problems</th>
<th>Interviewer Reading Problems</th>
<th>All Interviewer Reading Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>p&lt;.01</strong></td>
<td>-.001 (ns)</td>
<td>-.001*</td>
<td>-.002**</td>
<td>.0005 (ns)</td>
<td>.0002 (ns)</td>
</tr>
<tr>
<td><strong>p&lt;.01</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p&lt;.05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- H4a: Respondent fatigue: Respondents will show greater problems later in an interview than they do earlier. - NO
- H4b: Respondent learning: Respondents will show fewer problems later in an interview than they do earlier. - YES
- H5: Interviewer fatigue: Interviewers will show greater problems later in an interview than they do earlier. - NO
**H6a**: Interviewer reading times will decrease as the interview progresses. - YES

**H6b**: Response latencies will decrease as the interview progresses. - NO
  - **H6c**: Faster interviewer reading speed will be associated with faster respondent answers. - NO
  - **H6d**: Faster interviewer reading speed will be associated with more respondent difficulties. - NO

<table>
<thead>
<tr>
<th></th>
<th>Interviewer Reading Time</th>
<th>Response Latencies</th>
<th>Response Latencies</th>
<th>All Respondent Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Position</td>
<td>-.002**</td>
<td>-.0003</td>
<td>-.0005</td>
<td>-.001</td>
</tr>
<tr>
<td>Interviewer Reading Time</td>
<td></td>
<td></td>
<td>-.09**</td>
<td>.02**</td>
</tr>
</tbody>
</table>

**p<.01**
LIMITATIONS

- Did not use representative samples
- Laboratory setting for interview
- Behavior coding only captures verbalized problems
- Behavior coding – need to make inferences about what different behaviors mean
- Interviewers were matched to respondents on race/ethnicity and only a small number of interviewers in some strata
CONCLUSIONS AND FUTURE DIRECTIONS:

- **Conclusions:**
  - Question characteristics, interviewer and respondent behaviors
    - Question characteristics affect both respondent and interviewer behaviors
    - Interviewer behaviors affect respondent behaviors
    - The effect of question characteristics on respondent behaviors is not mediated by interviewer behaviors
  - Over time behavioral changes
    - Respondent learning
    - No apparent effects of fatigue
    - Faster interviewer reading, but no negative implications for respondent problems

- **Future directions**
  - Continued analysis of this data – rich data to test hypotheses about interactions between respondents and interviewers
  - Evidence of mechanisms?
<table>
<thead>
<tr>
<th>Options for Response Option Validity: (82% Valid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid response latency – I did not need to skip back or reread the question</td>
</tr>
<tr>
<td>Reread the question before I got to the response screen</td>
</tr>
<tr>
<td>Reread the question on the response screen</td>
</tr>
<tr>
<td>Reread the response options only</td>
</tr>
<tr>
<td>A probe or clarification was required</td>
</tr>
<tr>
<td>Skipped back to a previous question</td>
</tr>
<tr>
<td>Respondent answered before I finished reading the question</td>
</tr>
<tr>
<td>I struck the wrong key or waited too long to start/stop the timer</td>
</tr>
<tr>
<td>Something else went wrong (Other specify)</td>
</tr>
</tbody>
</table>
## Coding Scheme Used to Code Respondent Behaviors:

### Question Meaning Codes
- **(30)** Clarification (Unspecified):
- **(31)** Clarification (Construct/Statement):
- **(32)** Clarification (Construct/Question):
- **(33)** Clarification (Context):
- **(34)** Clarification (Time Frame):
- **(35)** Rewording (Question):
- **(36)** Respondent asks for repeat of question:
- **(37)** Difficulty using show card:
- **(38)** Clarification (Not enough information):

### Response Mapping Codes
- **(50)** Inadequate answer (General):
- **(51)** Clarification (Response format):
- **(52)** Respondent asks for repeat of response options:
- **(53)** Clarification (Response option meaning):
- **(55)** Rewording (Response options):
- **(60)** Imprecise response (General):
- **(61)** Imprecise response (Different response option):
- **(62)** Imprecise response (Inferred answer):
- **(63)** Imprecise response (Range):

### Memory Codes
- **(41)** Memory difficulty:
- **(42)** Making inferences:

### Response Mapping Codes
- **(01)** No problems identified:
- **(04)** Verbal reasoning:
- **(05)** Qualified answer (Uncertainty):

### Social Desirability Codes
- **(70)** Anonymity/confidentiality:
- **(80)** Refusal to answer (Privacy):
- **(81)** Refusal to answer (Other):
- **(82)** Refusal to answer (No reason):

### Other Respondent Codes
- **(90)** Respondent answers in different language:
- **(94)** Respondent laughs:
- **(97)** Other:
- **(98)** Don't know:
- **(99)** Missing data:
**Coding Scheme Used to Code Interviewer Behaviors:**

<table>
<thead>
<tr>
<th>Question reading problem codes</th>
<th>Other interviewer codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(110) Incomplete reading:</td>
<td>(131) Interviewer confirms/repeats respondent answer:</td>
</tr>
<tr>
<td>(111) Poor reading of question (Verbatim):</td>
<td></td>
</tr>
<tr>
<td>(112) Poor reading of question (Quality):</td>
<td>(132) Interviewer probes respondent answer:</td>
</tr>
<tr>
<td>(113) Interviewer self corrects:</td>
<td>(133) Interviewer assures respondent best answer is acceptable:</td>
</tr>
<tr>
<td>(114) Other question reading problem:</td>
<td></td>
</tr>
<tr>
<td>(117) Interviewer adds instruction prior to respondent answer:</td>
<td>(134) Interviewer offers respondent repeat of question:</td>
</tr>
<tr>
<td>(118) Interviewer omits show card instructions:</td>
<td></td>
</tr>
</tbody>
</table>

**Probing problem codes**

| (120) Poor probing: | (140) Interviewer provides respondent with answer: |
| (121) Missed probing: | (194) Interviewer laughs: |
| (122) Other probing problem: | (197) Other: |

**Other interviewer codes**

| (199) Missing data: | (101) No problems identified: |
DESCRIPTIVE STATISTICS FOR QUESTION CHARACTERISTICS

Question length: Number of sentences
    Mean=1.6 SD=.002

Text difficulty: Fleisch grade level
    Mean=10.6 SD=.01

Question response format
20%  Open-ended numeric response
    6%  Agree/disagree
    16% Yes/no
    17% Substantive categorical
    14% Unipolar Scale
    10% Fully Labeled Bipolar Scale (mdpt)
    2%  Fully Labeled Bipolar Scale (no mdpt)
    6%  Semantic differential – endpoints only
    1%  Semantic differential – endpts & mdpt
    2%  Initial branching item – with midpoint
    1%  Faces scale

Type of Judgment
    9%  Subjective
    89% Personal characteristics or behavior
    2%  Objective factual knowledge

Explicit DK or no opinion option? (4%)
Preceded by a DK filter question? (1%)
DK filter question? (1%)
Time qualified? (48%)
Was a show card used? (3%)
Designed to be difficult? (1%)
Proxy report? (18%)
Sensitive question or known to have SD connotations?
    Mean=.71 SD=.001

Level of abstraction
    Mean=.70 SD=.001