Comparing the Interviewer Variance Introduced by Standardized and Conversational Interviewing

Brady T. West\textsuperscript{1,2}, Frederick G. Conrad\textsuperscript{1,2}, Frauke Kreuter\textsuperscript{2,3,4} and Felicitas Mittereder\textsuperscript{1,3}

1. Survey Methodology Program, Institute for Social Research (ISR), University of Michigan
2. Joint Program in Survey Methodology, University of Maryland
3. Institute for Employment Research (IAB), Nuremberg, Germany
4. University of Mannheim, Germany

AAPOR 2015 Panel: Interviewer-Respondent Interactions
Acknowledgements

• Financial support for this research was provided by a grant from the National Science Foundation (SES-1324689) and the Institute for Employment Research (IAB) in Nuremberg, Germany

• We thank IAB staff (especially Daniela Hochfellner, Malte Schierholz, and Ulrich Thomsen) for their dedicated assistance with sampling and data access through the IAB-RDC at the University of Michigan

• We also thank the staff of infas in Bonn, Germany, for their dedicated assistance with all of the data collection operations and data quality checking
Research Question

• Standardized interviewing (SI) is widely used to ensure consistent administration of survey content and believed to minimize interviewer effects.

• A body of literature exists indicating that conversational interviewing (CI), designed to ensure respondent comprehension, can decrease response bias (e.g., Conrad and Schober, 2000, POQ); but critics wonder about an...

• **Open Question**: Does CI produce higher interviewer variance in survey responses than SI?
  • Uneven implementation, variance in wording, etc. may introduce more variance in responses across interviewers.
Study Design

• Original face-to-face data collection in 15 large geographic areas in Germany
• Simple random samples of 480 currently-employed adults drawn from each of the 15 areas
  • Adults had history of at least one unemployment spell
  • Samples drawn from government database (IEB) of official employment histories in each area
    • \( n = 7,200 \) in full sample; 4 interviewers per area
• The 4 interviewers in each area were each assigned 120 cases at random
  • Interpenetrated design; need to control for area effects
Study Design, cont’d

- Two interviewers in each area were rigorously trained in CI, and the other two were rigorously trained in SI
- Data Collection Period: April 2014 - October 2014
- Interviewers administered a 30-minute CAPI instrument
- The instrument included questions that we judged to require complex response processes, related to housing conditions, employment histories, and social networks
- Many questions were explicitly constructed to enable response validation using data on the IEB frame
- \( n = 1,850 \) respondents total (about 30 per interviewer)
Interviewer Training

• One full day of training in the two techniques (consistent with work of Conrad and Schober), including separate sessions for the two groups, conceptual clarification, testing, and role playing

• Audio recordings of interviews were consistently monitored by INFAS and the study team, and feedback was provided to interviewers on a weekly basis

• Initial analyses of audio recordings indicate that the two techniques were administered correctly and consistently (see upcoming AAPOR presentation by Mittereder et al., 10:15am, Sunday 5/17)
Survey Items

- **Household Questions** (e.g., number of rooms; number of residents; various recent expenses, such as moving)
- **Employment Questions** (e.g., any unemployment; hours worked per week; annual income)
- **Job Satisfaction and Commuting Preferences**
- **Work History in Past 20 Years** (e.g., number of times registered as unemployed)
- **Social Networking and Demographics**
- **Paradata** (e.g., total interview time, consent for record linkage, post-survey interviewer observations)
- **52 variables in total** for analysis purposes
Analytic Plan

- Multilevel linear, logistic, and ordinal models for each survey variable, with fixed effects of the CI technique and 14 of the 15 areas (necessary control!), and random interviewer effects

- Models allow the interviewer and residual variance components (for continuous items) to vary for the two groups; for group i, i = interviewer, j = respondent):

  \[ y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij} I[CI_i = 1] + u_{i1} I[CI_i = 1] + u_{i2} I[SI_i = 1] + \epsilon_{ij} \]

  \[ u_{i1} \sim N(0, \tau_{CI}^2), \quad u_{i2} \sim N(0, \tau_{SI}^2), \quad \epsilon_{ij} \sim N(0, \sigma_{CI}^2) \text{ if } CI_i = 1, \quad \epsilon_{ij} \sim N(0, \sigma_{SI}^2) \text{ if } SI_i = 1 \]

- Differences in variance components tested using methods outlined by West and Elliott (2014, *Survey Methodology*)
Criteria for “Better” Data Quality

• **Criterion 1:** Intra-interviewer correlations (ICCs)
  • A conventional criterion: Which group has a smaller ICC?

• **Criterion 2:** Ratios of Variance Components (less Variance “better”)
  • Ratio of CI interviewer variance (IV) to IV for SI
  • Ratio of CI residual variance (RV) to RV for SI (for continuous variables)
  • Are the components significantly different from each other?

• **Criterion 3:** Differences in Means / Proportions (Bias)
  • Is the fixed effect of CI significant? Use standardized test statistic...
  • If so, which group has a higher mean? Does it suggest higher accuracy?
  • Example: For number of rooms, YES, given mostly inclusive definition (e.g., include bathrooms); for recent home improvement expenses, NO, given mostly exclusive definition (e.g., exclude maintenance)
  • Ideally: examine mean response errors using IEB validation data

• We examined all three criteria for each survey item, and mixed results across the three criteria were entirely possible
Differences: Household Variables (3 of 7)

Mixed result for number of rooms: significant positive effect of CI (good, given inclusive definition), but more IWER variance (bad).

Significant negative fixed effects of CI, and no differences in variance. CI “better”: conceptual definitions exclusive.
Differences: Employment (9 of 16)

Criterion 1

- ANYMARG
- ANINCEX
- ANNING
- MNTHINC
- LNGTHEMP
- HRSMARG
- HOURSWK
- NUMEMPS
- TNUMEMPS

Criterion 2

- ANYMARG
- ANINCEX
- ANNING
- MNTHINC
- LNGTHEMP
- HRSMARG
- HOURSWK
- NUMEMPS
- TNUMEMPS

Criterion 3

- ANYMARG
- ANINCEX
- ANNING
- MNTHINC
- LNGTHEMP
- HRSMARG
- HOURSWK
- NUMEMPS
- TNUMEMPS

Employment Questions

- TNUMEMPS
- NUMEMPS
- HOURSWK
- HRSMARG
- LNGTHEMP
- MNTHINC
- ANNING
- ANINCEX
- ANYMARG

Employment Questions

- TNUMEMPS
- NUMEMPS
- HOURSWK
- HRSMARG
- LNGTHEMP
- MNTHINC
- ANNING
- ANINCEX
- ANYMARG

Employment Questions

- TNUMEMPS
- NUMEMPS
- HOURSWK
- HRSMARG
- LNGTHEMP
- MNTHINC
- ANNING
- ANINCEX
- ANYMARG

Mixed: CI lower probability of reporting marginal employment (exclusive), but more IWER variance

CI increases probability of exact income report!

Mixed: CI higher reports of employee counts (good); more residual variance

CI better for length of employment & marginal job hours (exclusive def'ns)

SI “better” for income reports (less variance)

SI “better” for regular job hours

Mixed: CI/CI IV Rat, * p<.05

CI/CI RV Rat, * p<.05

CI/CI TotVar Ratio

AAPOR 2015 Panel: Interviewer-Respondent Interactions

11
Differences: Work History (2 of 4)

• For reports of the longest uninterrupted period of gainful employment in the past 20 years, CI produced substantially higher interviewer variance ($p < 0.01$):
  – ICC for CI: 0.065, ICC for SI: 0.001
  – Ambiguity in what should be considered an interruption; no significant fixed effect of CI

• For # of times registered as unemployed, SI produced substantially higher within-interviewer variance
  – Possible benefits of CI...
  – Whether or not this is bad requires validation data!
Differences: Social Networks (3 of 9)

• For reports of counts of close friends outside of the house, mixed findings:
  – CI yielded marginally higher interviewer variance
  – ICC for CI = 0.026, ICC for SI = 0.001; no fixed CI effect
  – SI yielded significantly higher within-interviewer variance \( (p < 0.05) \) and higher total variance

• CI reduces the odds of saying that you belong to a church (suggestive of less social desirability?)

• CI increases the odds of saying that you use a social networking web site (less social desirability?)
Differences: Paradata (4 of 5)

**Criterion 1**
- QUALUND
- INTTIME
- QUALINF
- QUALUND

**Ratio**
- CI/SI IV Rat, * p<.05
- CI/SI RV Rat, ^ p<.05
- CI Eff/SE (* p<=.05)

- CI consistently yields post-survey interviewer observations that indicate HIGHER data quality.

- Both ICCs for interview time are extremely high!

**Criterion 2**
- QUALUND
- INTTIME(*)
- QUALINF(*)
- QUALUND(*)

**Ratio**
- CI Eff/SE (* p<=.05)
- CI/SI TotVar Ratio

- CI produces slightly longer interviews, and more within-interviewer variance in length.

**Criterion 3**
- QUALUND(*)
- INTTIME
- QUALINF
- QUALUND

**Ratio**
- CI/SI IV Rat, * p<.05
- CI/SI RV Rat, ^ p<.05

- Why would CI reduce the odds of consenting to audio recording?

CI consistently yields post-survey interviewer observations that indicate HIGHER data quality.

Both ICCs for interview time are extremely high!

CI produces slightly longer interviews, and more within-interviewer variance in length.

Why would CI reduce the odds of consenting to audio recording?
Additional Findings

• No significant results related to 4 principal components measuring job satisfaction

• No significant results related to 3 demographic items (expected)

• For commuting, no significant results for 3 items, but one difference for a fourth:
  – Reported distances to the nearest train station had higher within-interviewer variance for CI ($p < 0.01$), favoring SI
Summary of Results

• Reminder: 52 analysis variables in total
• CI “Better”: 10 variables
  – Recent household expenses
  – Length of employment at current job, weekly hours in marginal jobs, exact income reports, times registered as unemployed
  – Reduced likelihood of socially desirable social network responses
  – Interviewer observations from the CI group suggest that data of higher quality were provided by respondents
Summary of Results, cont’d

• SI “Better”: 7 variables
  – Less variance (both between and within interviewers) in reports of monthly and annual income
  – Less variance in reports of regular weekly job hours and longest periods of gainful employment
  – Smaller within-interviewer variance in reported distances
  – Shorter interviews with less within-interviewer variance in times; increased odds of consenting to audio recording

• Mixed Findings: 5 variables

• No Differences: 30 variables
Implications of Results

• The final “tally” suggests modest overall benefits of CI relative to SI, but there are some mixed results...
  – For the mixed results, is bias or variance a larger concern?
  – **Example:** The two groups had similar total variance for number of rooms, suggesting that CI may be better for reducing bias (if the higher ICC does not also impact MSE)

• Certain techniques may be better for certain types of questions; switch approaches within interviews?

• Organizations should not subscribe exclusively to one technique or another!
Next Steps

• Analyses of the administrative data (for validation), to confirm these initial findings (work in progress...)
  – 2014 employment data not available until Nov. 2015
  – Early work suggests that CI $\rightarrow$ higher response accuracy
• Requires working with IEB administrative data, which is not easy (recorded in unique spells...)
• Decomposition of interviewer variance for each of the two techniques into measurement error variance and nonresponse error variance
• Validation work to be presented at JSM in August
Thank You!

• Please direct any questions to bwest@umich.edu.
2015 International Total Survey Error Conference

TSE15

Improving Data Quality in the Era of Big Data

Save the Date

September 19-22, 2015
Renaissance Baltimore Harborplace Hotel
Baltimore, Maryland, USA

www.tse15.org