University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective

Sociology, Department of

2-26-2019

nvestigating the Utility of Interviewer Observations on the Survey Response Process

Frauke Kreuter
University of Maryland/University of Mannheim/Institute for Employment Research, fkreuter@umd.edu

Brady West
University of Michigan, bwest@umich.edu

Ting Yan
Westat, tingyan@westat.com

Michael Josten
Conflict Management Consulting

Heather Schroeder
University of Michigan, schroedh@umich.edu

Follow this and additional works at: http://digitalcommons.unl.edu/sociw

 Part of the Quantitative, Qualitative, Comparative, and Historical Methodologies Commons

Kreuter, Frauke; West, Brady; Yan, Ting; Josten, Michael; and Schroeder, Heather, "nvestigating the Utility of Interviewer Observations on the Survey Response Process" (2019). *2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective*. 18.
http://digitalcommons.unl.edu/sociw/18

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in 2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Utility of Interviewer Observations on the Survey Response Process

Michael Josten Conflict Management Consulting
Heather Schroeder University of Michigan
Brady West University of Michigan
Ting Yan Westat
Frauke Kreuter University of Maryland/Mannheim & IAB

The National Survey of Family Growth (NSFG) is conducted by the Centers for Disease Control and Prevention's (CDC's) National Center for Health Statistics (NCHS), under contract # 200-2010-33976 with University of Michigan's Institute for Social Research with funding from several agencies of the U.S. Department of Health and Human Services, including CDC/NCHS, the National Institute of Child Health and Human Development (NICHD), the Office of Population Affairs (OPA), and others listed on the NSFG webpage (see http://www.cdc.gov/nchs/nsfg/). The views expressed here do not represent those of NCHS nor the other funding agencies.

Research Questions

Can interviewers effectively

- **rate the respondent's performance** in surveys?
- be used to derive a **meaningful single indicator of response quality**?
- indicate **where the quality breaks down** but pointing to specific steps in the cognitive response process (comprehension/retrieval/judgement/editing)?

Past Findings – Use of Interviewer Ratings

- Bennett (1948) implemented questions about the quality of respondent's answers addressed to the interviewer as an **instrument against interviewer cheating**.
- Fisk, G. (1950) evaluation of interviewer observations of the survey response process focused on the **variation of interviewer ratings** of interest **between interviewers**
- Feldman, Hyman, and Hart (1951) multiple interviewer evaluations on the respondent's behavior throughout the interview -> **analyze interviewer variance and interviewer influence**

- Later studies found respondents who received **positive or favorable ratings** tended to provide data of better quality in terms of a variety of indicators.
  - less missing data (Tarnai, J., and Paxson, M.C. 2005; Antoun, C. 2012),
  - less measurement error (Peytchev, A. and Peytcheva, E. 2007),
  - more consistent reports (Antoun, C. 2012), and
  - more codeable answers to open-ended questions (Tarnai, J., and Paxson, M.C. 2005).

Data - NSFG

- Interviewer observation data from the National Survey of Family Growth, a national survey of sexual and reproductive habits in 15-49 year olds.
- 60 minutes survey with two sections, an in-person computer assisted personal interview (CAPI) and an audio computer-assisted self-interview (ACASI).
- 15,820 interviews over 12 quarters of data collection, 1/2016 – 12/2018.
- 30 post-interview observations, including details about the environment, the respondent's response behaviors, and respondent mood. Here 22 of the 30 observations used to map the survey response process.
- Each observation was classified into one of the four stages of the survey response process. (n=52 excluded due to missingness -- final sample size of 15,768).

Observations – Comprehension

- Comprehension: include four indicators (yes/ no) of **distractions**:
  - television on,
  - respondent received phone calls,
  - children present and need attention,
  - and other distractions.
- Comprehension of the **ACASI**
  - how much help the respondent needed from the interviewer (none/ a little/ a lot/ or interviewer administered),
  - the respondent's use of headphones (at least some of the time/ never),
  - the respondent's use of text and audio (text only/ text and audio/ audio only/ don't know),
  - and what support was used to hold the laptop (table/ lap/ other).
  - Difficulty using the CAPI application (any/ none)
  - Interviewer's opinion of respondent attentiveness (not at all, some/ very)

Observations – Retrieval/Estimation/Editing

- Retrieval
  - Was the respondent upset (yes/ no),
  - Tired (yes/ no),
  - Tow did they act during the interview (hostile, neutral/ friendly).
- Judgement or estimation process was an overall indication of the quality of information provided by the R (excellent/ good/ fair-poor).
- Mapping and editing process
  - Seating arrangement (next to respondent facing the same way/ next to respondent at a right angle/ across from the respondent/ other)
  - Presence of other persons within hearing range (no one else present/ 1+ people present, not able to hear/ 1+ people present, able to hear part of IW/ 1+ present able to hear whole IW).
  - Measures of the respondent's ability to see the computer screen during the CAPI section (yes, all questions/ most, not all questions/ a few questions/ none),

Observations – cont'd

- Interviewer's ability to see the computer screen during the ACASI section (yes/ no)
- Interviewer's mood
  (happy/ neutral, sad, unhappy) were also captured.
- Location (on the respondent's property/ in the interviewer's car/ in another public place)
- Atmosphere
  (chaotic, noisy/ some interruptions/ ideal- quiet and calm)
- Language of the interview (English/ Spanish/ both)

Method

- Six latent class analysis models that each included the 22 categorical interviewer observations.
- Each model used a different number of latent class groups ranging from 2 to 7.
- Fit statistics (including log likelihood, g-squared, AIC, BIC, and adjusted BIC) and entropy (higher entropy indicators better class separation) compared across the 6 models.
- Focus on the model with 7 latent classes that minimizes the model fit statistics and has an entropy estimate of 0.89.

Results

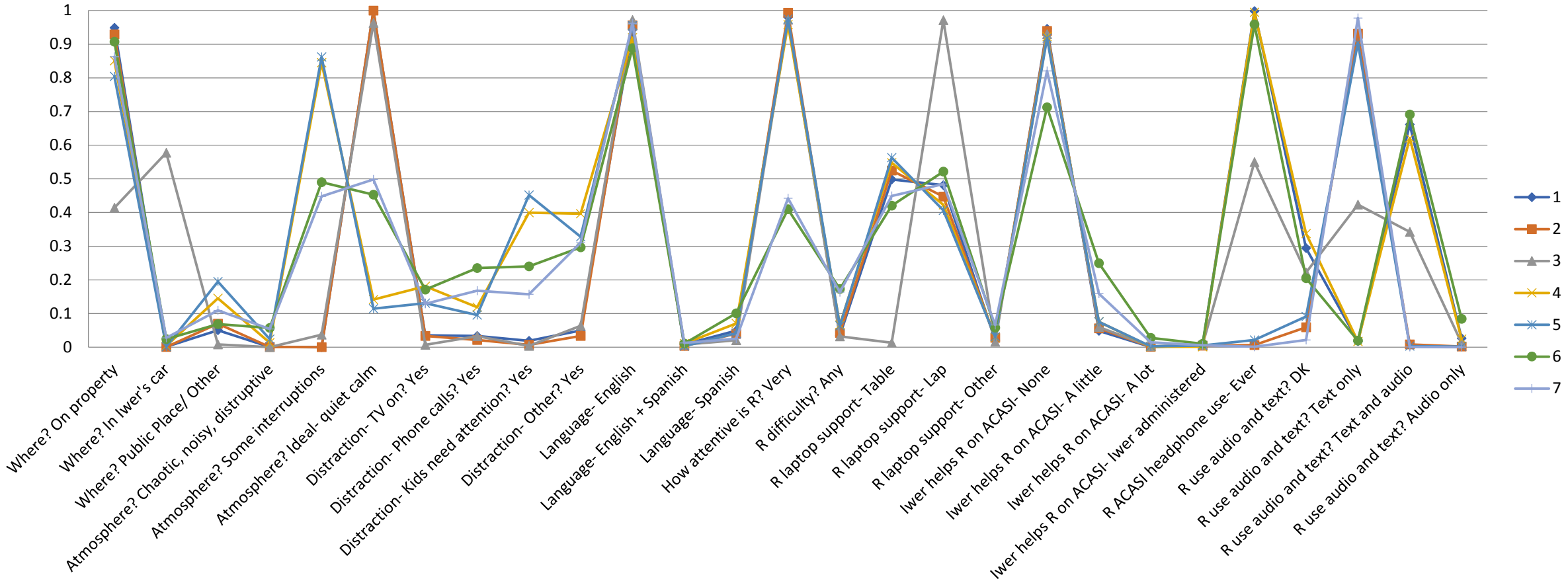
- 27% in class 1, private, text ACASI, high quality
- 36% into class 2 most private, high quality

- 6% in class 3 distractions with kids, text ACASI, no R problems
- 8% in class 4 distractions, no kids, R problems, iwer assist, iwer unhappy, headphones in ACASI, low quality
- 10% in class 5 distractions with kids, headphones in ACASI, no R problems

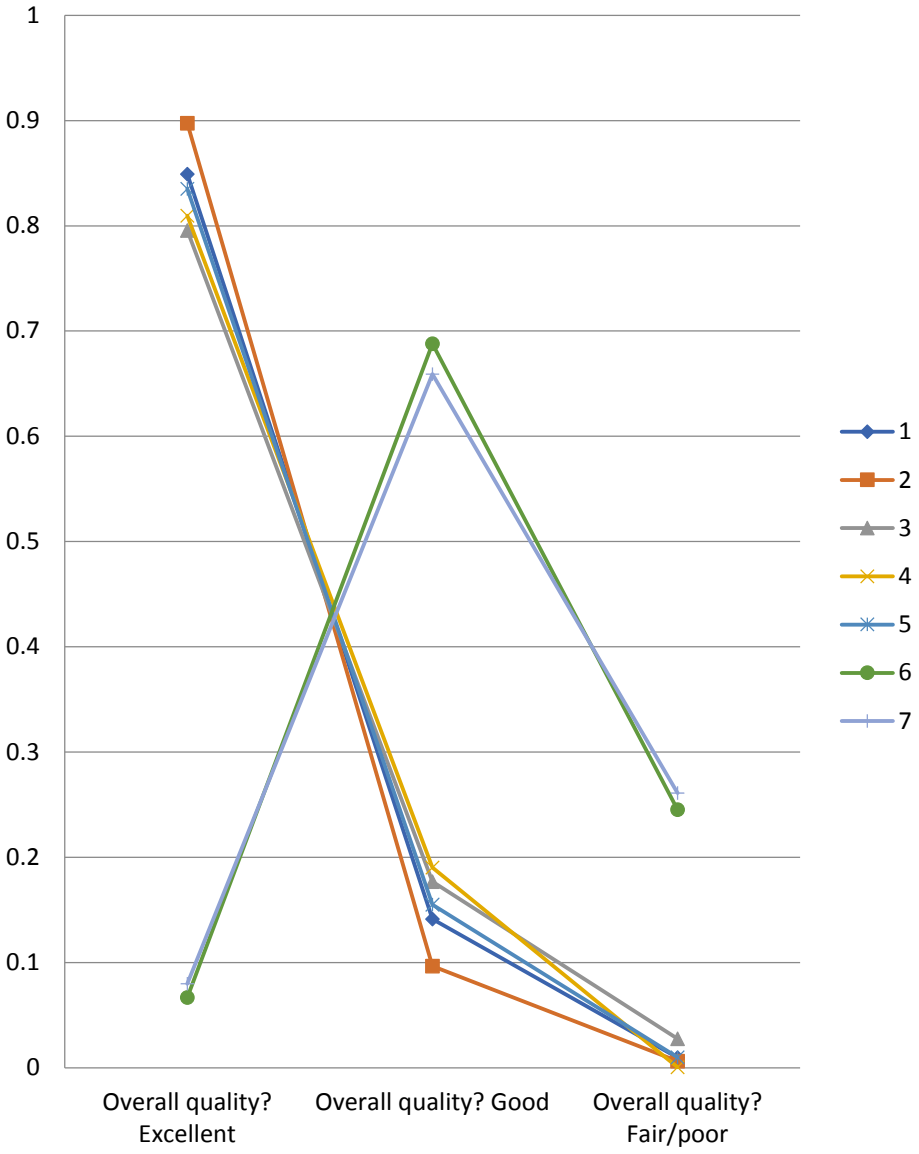
---

- 8% in class 6 car interviews, sometimes headphones in ACASI
- 5% in class 7 worst quality across board, no privacy, iwer unhappy

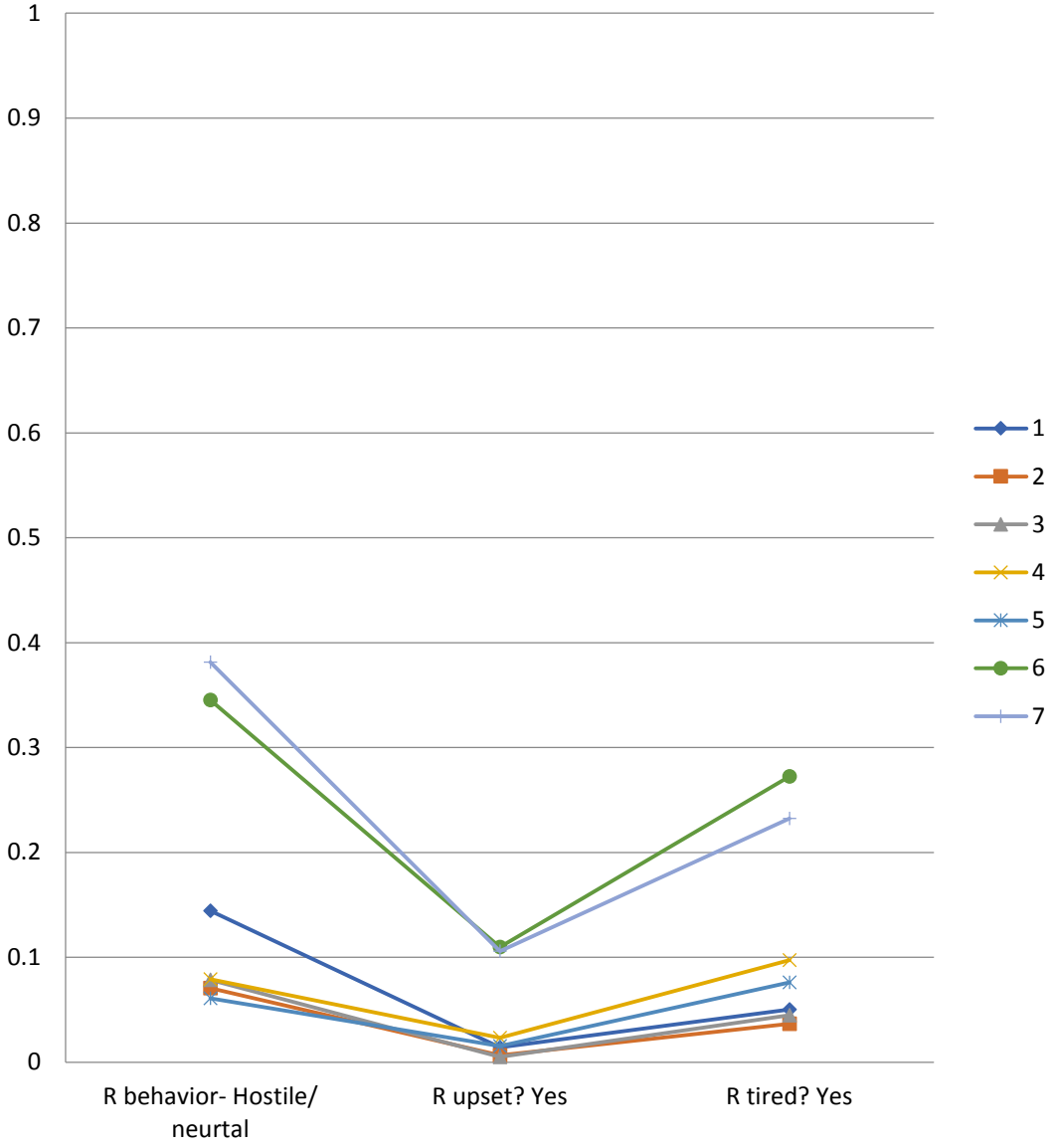
Comprehension



Judgement Stage



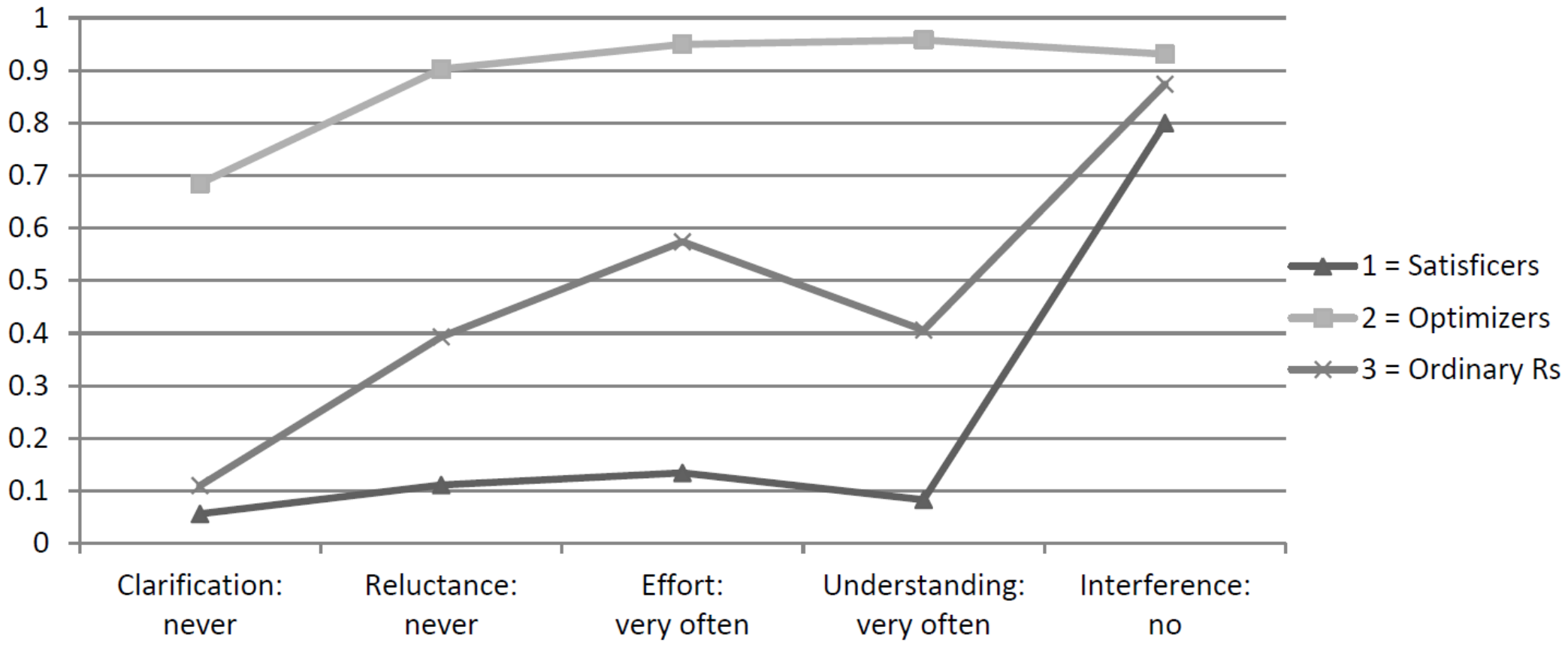
Retrieval Stage



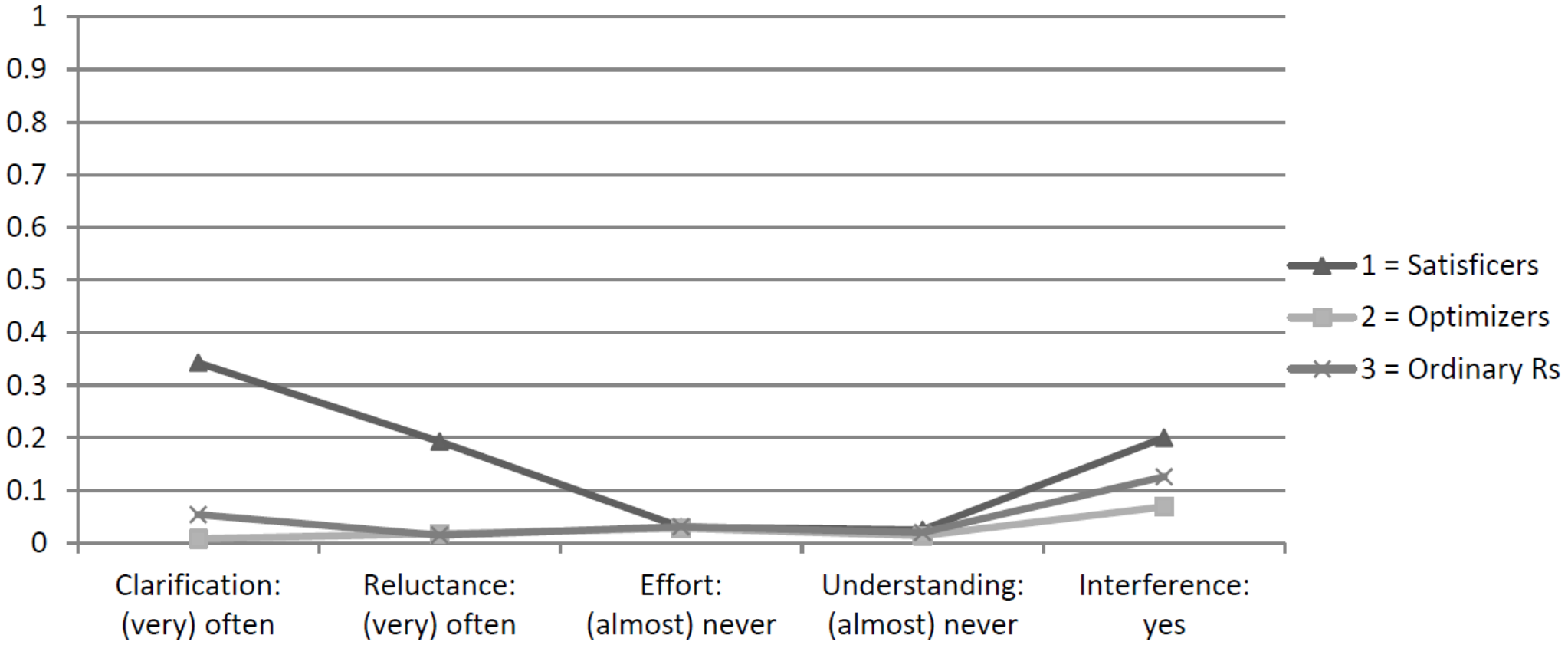
Data – ESS

- 5th wave of ESS used for this analysis (mostly CAPI)
- 15,820 interviews over 12 quarters of data collection, 1/2016 – 12/2018.
- **Interviewer ratings** (5 point scale; never – very often; don't know):
  - Understanding of the question
  - Clarification of any questions
  - Reluctance in answering
  - Answering with best effort
  - Presence of others with interference
- **Quality indicators**
  - non differentiation (at least 4 same items in matrix question),
  - extreme and middle answers,
  - acquiescence (percentage agreeing in a set of 23 attitudinal items)
  - internal consistency
  - duration

Conditional Probabilities – highest quality



Conditional Probabilities – poorest quality



LCA Rating Results – Quality Indicators

	Freq.	%
Class 1 - Satisficers	1845	11.67
Class 2 - Optimizers	9046	57.20
Class 3 - Ordinary Rs	4923	31.11

	NR-rate	Non-diff.	Extreme	Acqui.	Incon.	Length
Class 1 - Satisficers	$6.60^{2,3}$	$12.94^{2,3}$	$24.81^{2,3}$	54.84	$3.85^{2,3}$	17.32
Class 2 - Optimizers	2.97	8.30	23.85	$59.11^{1,3}$	2.86	17.08
Class 3 - Ordinary Rs	4.08^{2}	10.56^{2}	23.51	57.51^{1}	2.60	17.50^{2}
Total	3.74	9.54	23.86	58.12	2.90	17.24

Numbers are percentages, except for interview length which is measured in seconds per item.

Superscript numbers indicate statistically significant differences ($p < 0.05$) against respective class.

Discussion

- Interviewer observations have the potential to predict data quality and/or response differences.
- Given cost of interviewer observations sensible to investigate
  - which observations are more useful for which types of measures,
  - whether the observations can consistently predict data quality.
- Systematic analysis of observations across different surveys needed
- Observations only recorded at the end – possible that distractions or other events occurred but not in a meaningful way for the entire interview. More mid-stream recording?

Thank you!

fkreuter@umd.edu