Associations Between Interactional Indicators of Problematic Questions and Systems for Coding Question Characteristics

Jennifer Dykema
University of Wisconsin Survey Center
University of Wisconsin-Madison

Nora Cate Schaeffer
Department of Sociology
University of Wisconsin Survey Center
University of Wisconsin-Madison

Dana Garbarski
Center for Women's Health and Health Disparities Research
University of Wisconsin-Madison

Annual Meeting of the American Association for Public Opinion Research
Boston, MA
17 May 2013

Background and research objectives

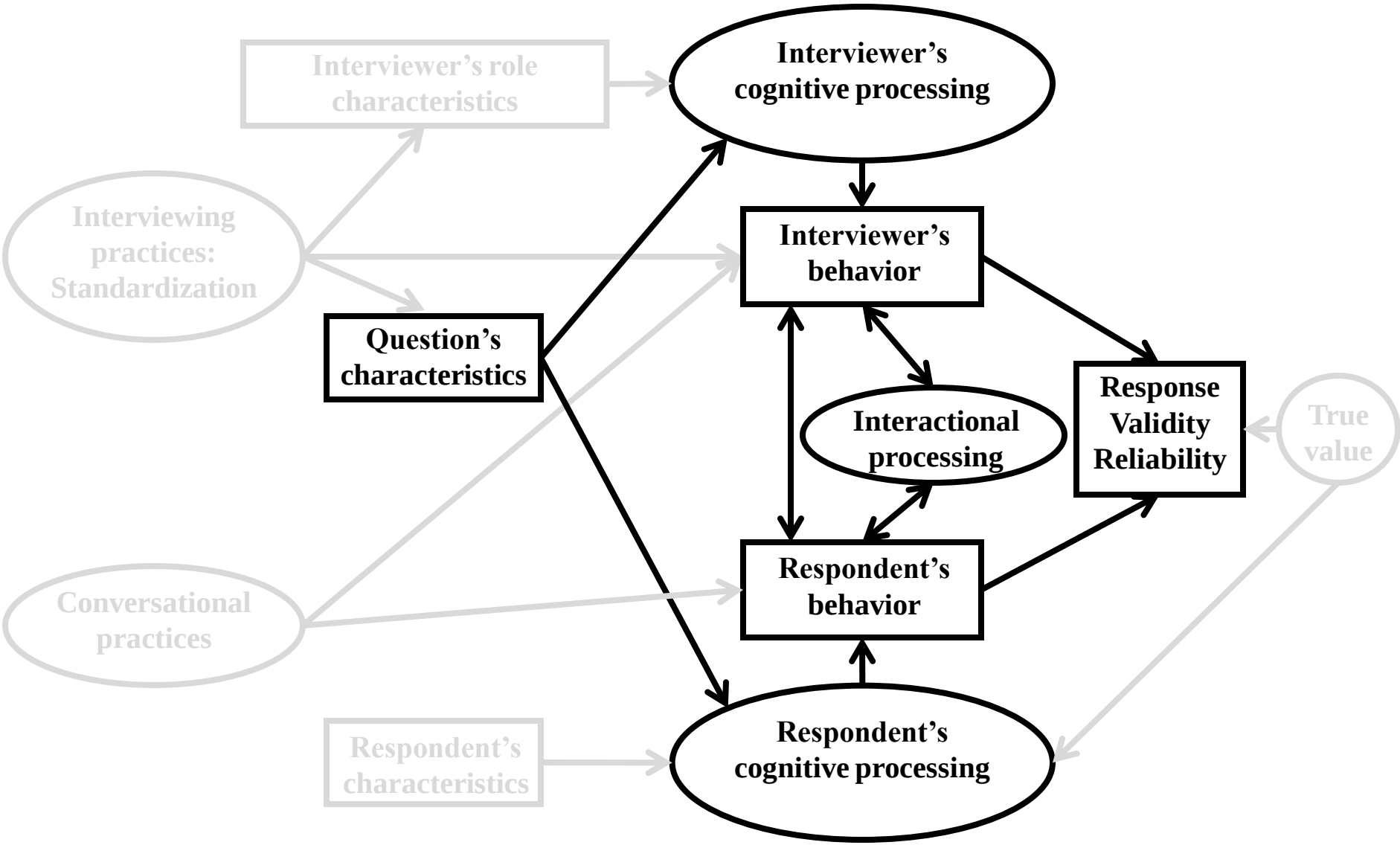
- Question writers often focus on question characteristics
  - e.g., length, difficulty, response format
- Recommendations for writing questions are
  - formulated around question characteristics
  - based on research (beliefs) about impact of question characteristics on outcomes

- Know a lot about effects of some question characteristics on data quality
- Still developing a comprehensive typology in which
  - question characteristics are cataloged
  - effects on INTs' and Rs' processing are understood
  - effects on data quality are documented

Research questions

- What are (some of) the approaches used to conceptualize, measure, and code question characteristics?
- How do they differ?
- Which succeed in predicting interviewer-respondent interactional behaviors during the interview?

Interactional Model of Question-Answer Sequence



Approaches and summary of dimensions

- Approaches
  - Question length
  - Readability
  - Question Understanding Aid (QUAID)
  - Problem Classification Coding Scheme (CCS)
  - Question Appraisal System (QAS)
  - Survey Quality Predictor (SQP)
- Dimensions (summary)
  - What it is
  - How scored
  - Focus
  - Goal
  - Scope
  - Reliability (of our internal coding)

Approaches and dimensions: Question length

Dimension	Description
What it is	Simple count of total number of words in question's text
How scored	Computer: Count words in MS Word
Focus	Language and grammar
Goal	Find problems (longer question, more problematic)
Scope	Narrow No allowance for difficulty, other features
Reliability	High

Approaches and dimensions: Readability

Dimension	Description
What it is	Flesch-Kincaid Grade Level score indicates comprehension difficulty in a passage of text (question)
How scored	Computer: Feed text into MS Word, Flesh 2.0 (http://flesh.sourceforge.net/)
Focus	Language and grammar
Goal	Find problems (higher grade level, harder to comprehend)
Scope	Less narrow; Formula takes into account ratio of words to sentences and syllables to words
Reliability	High

Approaches and dimensions: Question Understanding Aid (QUAID) (Graesser et al. 2006)

Dimension	Description
What it is	Tool to evaluate questions on comprehension difficulty: unfamiliar technical terms, vague or imprecise relative terms, vague or ambiguous noun phrases, complex syntax, and working memory overload
How scored	Computer: Enter text, QUAID returns list of problems (http://mnemosyne.csl.psyc.memphis.edu/QUAID/quaidindex.html) We tallied number of categories with problems
Focus	Language and grammar
Goal	Find problems (higher counts, more comprehension difficulties)
Scope	Less narrow; Considers multiple categories related to comprehension
Reliability	High

Approaches and dimensions: Problem Classification Coding Scheme (CCS) (Forsyth et al. 2004)

Dimension	Description
What it is	Scheme for coding 28 problems; Problems grouped under the 4-stage question-answer model comprehension, retrieval, judgment, response E.g., “Comprehension and communication,” “Question content,” “Vague topic/term”
How scored	Trained coder codes question, number of problems tallied
Focus	Demands at different stages of the question-answer process
Goal	Find problems (higher counts, more problems)
Scope	Relatively comprehensive: Attempts an overall evaluation of cognitive issues and includes categories for interviewer-related problems.
Reliability	Moderate

Approaches and dimensions: Question Appraisal System (QAS) (Willis 2005)

Dimension	Description
What it is	Scheme for coding 27 problems; Focus on question characteristics likely to cause problems: categories for reading, instructions, clarity, assumptions, knowledge, sensitivity, response categories, and other; E.g., "Clarity," "Vague: There are multiple ways to interpret the question or to decide what is to be included or excluded."
How scored	Trained coder codes question, number of problems tallied
Focus	Finding problems with questions or answers
Goal	Find problems (higher counts, more problems)
Scope	Relatively comprehensive: Attempts an overall evaluation of cognitive issues and includes categories for interviewer-related problems.
Reliability	Moderate

Approaches and dimensions: Survey Quality Predictor (SQP) (Saris and Gallhofer)

Dimension	Description
What it is	Tool for coding language, structure, content, administration; Obtain a quality predictor based on previously conducted analysis of question characteristics using MTMM data analysis; E.g., "Response scale," "Number of categories" and "Labels of categories"
How scored	Human coder codes question characteristics in SQP 2.0 (http://www.sqp.nl/), program outputs scores for reliability, validity, and quality
Focus	Predicting quality
Goal	Measure quality (higher score, higher data quality)
Scope	Very comprehensive: Depending on the content and structure of question being evaluated, close to 50 characteristics may be coded
Reliability	Moderate

Effects of Question Characteristics on Interactional Outcomes

Survey data: The Wisconsin Longitudinal Study

- 1/3 random sample of Wisconsin high school class of 1957
- Telephone interviews digitally recorded in 2004
- 355 cases randomly sampled
- Analyze question characteristics using 23 questions in the Health module
  - *In general, would you say your health is excellent, very good, good, fair, or poor?*
  - *Have you been able to bend, lift, jump and run without difficulty and without help or equipment of any kind?*

Behavioral outcomes: Interaction coding data

- Behavioral outcomes from interviewer-respondent interactions
- Interviews transcribed, coded in Sequence Viewer (Dijkstra)
- Elaborate coding scheme: Over 100 behaviors
  - Ex: pauses, tokens, uncodable answers, etc.
  - Small subset analyzed here
- Question-answer sequence
  - Unit of analysis
  - Starts with reading of the survey question by INT, ends with the last utterance spoken by INT or R before INT reads next question
  - Made up of behaviors
    - Each utterance spoken by INT or R is coded
  - 8150 question-answer sequences

Behavioral outcomes: Dependent Variables

- Interviewers
  - Question-reading accuracy
    - exact versus any changes
  - Tokens (any)
    - e.g. “well,” “um,” “oh,” “er”
    - terms or phrases with a neutral connotation linked to processing difficulties
- Respondents
  - Index of “problem” behaviors
    - uncodable answers, qualified answers, etc.
  - Tokens (any)
  - Ask questions/seek clarification (any)

Methods

- Question characteristics: Independent Variables
  - Larger values = relatively more problematic question
    - Question length, Flesch, QUAID, CCS, QAS
  - Larger values = relatively less problematic question
    - SQP
  - Use standardized scores (z-scores) in the analysis
- Modeling
  - Data have a complicated multilevel structure
    - Rs nested within INTs; Qs crossed by Rs and INTs
  - Use mixed effects logistic regression models
  - Include random effects for INTs, Rs within INTs, Qs (after fixed effects), and INTs crossed by Qs
  - Results from BIVARIATE models

Odds ratios from bivariate mixed effects logistic regression equations of interviewer behaviors on question characteristics

	Exact Reading	Token (Any)
Fixed Effects	**Odds Ratios**	**Odds Ratios**
Question Characteristics		
Question length	**0.37*** **	**1.29** **
Flesch grade level	0.80	**1.23***
CCS	**0.53** **	**1.31** **
QAS	**0.65+**	1.08
QUAID	**0.59***	1.18
SQP	**1.82** **	0.91

Odds ratios from bivariate mixed effects logistic regression equations of respondent behaviors on question characteristics

	Problems (Any)	Token (Any)	Qs (Any)
Fixed Effects	**Odds Ratio**	**Odds Ratio**	**Odds Ratio**
Question Characteristics			
Question length	**2.62*** **	**1.58** **	**1.60***
Flesch grade level	1.26	1.03	1.21
CCS	**1.75***	**1.50+**	**1.84** **
QAS	1.26	1.09	**1.41+**
QUAID	**2.29*** **	**1.60** **	**1.50***
SQP	0.74	0.90	0.88

Summary of results, limitations, and future directions

- Summary of results
  - Surprisingly direction of effects for all of the coding approaches were in the predicted direction for interviewer question asking and significant for most
  - The most consistent predictors of the respondent problem behaviors were question length, CCS, and QUAID
  - Not drawing the conclusion that long questions → poor quality data
    - Questions in this study that were long were also complex
    - Future work could test the interaction of question length with other characteristics

Summary of results, limitations, and future directions

- Limitations
  - Coding approaches are not independent (e.g., most of the approaches code for question length in some way)
  - Questions were not randomly sampled from a population of questions with many different characteristics
    - Questions were primarily yes/no type questions about health
  - Limited number of interviewer, respondent, and interactional behaviors examined
  - Implicitly assume that behavioral measures are associated with poorer quality data

Summary of results, limitations, and future directions

- Future directions
  - Examine question characteristics and coding approaches with a bank of questions with more varied characteristics
  - Modeling building: experience and cognitive ability
  - Examine other approaches for coding question characteristics
    - Our system
      - More codes specific to interviewers
      - More detailed specifications for coding questions to maximize reliability
  - Incorporate measures of validity and reliability of survey responses as outcomes to predict

Thank You!

For copies of this presentation or more information, contact:

Jennifer Dykema
dykema@ssc.wisc.edu

Please visit us at:
www.uwsc.wisc.edu