Using Behavior Coding to Evaluate Survey Questions

Jack Fowler

Center for Survey Research

Umass Boston

What should question-answer process look like?

- 1. I reads question exactly as worded
- 2. R understands question as intended
- 3. R retrieves information needed to answer
- 4. R puts answer in form required and tells interviewer

Premises of behavior coding

- 1. Deviations from this ideal may reflect problems that are threats to validity of data

- 2. The wording of a question often is the direct cause of these problems

- 3. The presence of problems can often be inferred from the behavior of interviewers or respondents

A little history

- Charlie Cannell et al were trying to understand what was going on when respondents were doing a good and poor job of reporting health events in the NHIS
- Created an observation scheme
  - Sampled questions
  - Ratings (anxious, bored) were used for some
  - Others were behavior coded (read question as worded, etc.)

Later adaptations

- Used as monitoring scheme to supervise interviewers

- Also used in studies of how training and supervision affected interviewer behavior

A serendipitous finding

- When looking at how training affected interviewer performance, investigators found that certain questions were routinely causing problems, regardless of how much training interviewers had

- This led to wondering if behavior coding could be used to flag problem questions

An anecdote about funding question-evaluation studies

- Applied to AHRQ (then NCHSR) for funds to study behavior coding as a question evaluation technique

- Reviewers panned the idea

- Charlie Cannell got ticked off at what he thought were really uninformed (stupid) reviews

- We appealed the ratings, asked for better reviewers, and we got funded

How to do behavior coding

- Can be done live using an observer
- Mainly done by recording interview and then having trained coders listen to recordings

- NOTE: Experience shows that almost everyone who is willing to be interviewed agrees to be tape recorded

What to code

- Unit of observation is usually the question

- The core focus of most coding is to count deviations from an ideal question and answer process

Question reading

- Read exactly as worded
- Minor changes
- Major changes
- Interrupted

Question answering

- Adequate (codable) answer given

- Inadequate (uncodable) answer given

- Qualified answer ("I think" "It might be..")

- Refusals and "don't knows"

Other aspects of R behavior

- Asks for clarification

- Asks for all or part of question to be repeated

Sample output for each question

- Often results are reported like this:
  - % read exactly as worded
  - % interrupted reading
  - % asked for clarification
  - % gave inadequate answer

What do results tell us? (Some examples)

- 1. Interruptions often occur when there is dangling material in the question, such as a definition after the question.

- How many children do you have? Do not include step children.

What do results mean?

- 2. Inadequate answers often occur when it is unclear how to answer the question.

- When did you move to New York?

- Unclear whether interviewer wants date, years ago, or stage in life cycle—and how precise it has to be.

What do results mean?

- 3. Requests for clarification often mean there is an unclear term or concept.

- When did you move to New York?

- Does that mean the city, the metropolitan area or the state?

Bases for interpreting results

- 1. We have some generalizations that have been pretty well established, such as the preceding

- 2. Cognitive testing may generate hypotheses that permit interpreting results

- 3. Feedback from interviewers or coders

- 4. Listening to the exchanges that led to the behaviors in question

Behavior Coding Results are Reliable

- Two interviewing staffs used same instrument in parallel—interviews were behavior coded

- When the rates of behaviors per question (e.g. reading as worded, inadequate answers) were tabulated and correlated for two staffs, correlations ranged from .6 to .8

Effects on Data

- One of the clearest findings is that the higher the rate at which interviewers have to probe to get an adequate answer, the higher the interviewer-related error

More effects on data

- There is evidence that qualified answers and response latency are related to the "accuracy" of responses

More effects on data

- There are examples in which unclear terms were inferred from the behavior coding results

- Revised questions "improved" the behavior coding results and produced estimates deemed likely to be more accurate

Strengths of behavior coding as question evaluation method

- 1. Low cost—easily integrated into pretest
- 2. Evaluation is of how questions work under realistic data collection conditions
- 3. Results are reliable
- 4. Results are objective—i.e. not dependent on an individual's subjective assessment
- 5. Results are quantitative

And for Q Bank?

- Behavior coding results would seem to be particularly well suited to be routinely included in Q Bank
  - Because results are quantitative
  - And because they are reliable

  They are ideally suited to facilitate comparisons of questions that were used in different surveys

Weaknesses of Behavior Coding

- 1. Sometimes hard to diagnose reason for results and how to fix it

- 2. Can't always tell if an observed problem actually affects data

- 3. Some question problems (such as comprehension) do not show up in behavior coding

Place of behavior coding in question evaluation

- It is not a substitute for expert review

- It is not a substitute for cognitive testing

- However, because it is quantitative, objective, reliable, easy to do, and provides data under realistic conditions, it is a highly valuable complement to those two question evaluation approaches

THANK YOU