Questionnaire Design for Social Surveys

Week 5 Testing Questionnaires

5.1. Qualitative Pretesting Techniques

Pretesting Techniques: Expert Reviews and Focus Groups

Why pre-testing / testing?
• Identify and reduce
  - specification error
    - decomposition
  - operationalization error
    - questions not measuring construct
    - variability across respondents/interviews/times
  - measurement error
    - question characteristics leading to response error
    - respondent processing issues leading to response error
    - interviewer characteristics/behaviors leading to response error
    - instrument features leading to response error

Tools for Testing Survey Questions
• Expert Review
• Focus Groups
• Cognitive Interviews
• Behavior Coding
• Statistical Models
  - Latent Class Analysis
  - Structural Equation modeling
• Field Test
• Other

Expert Reviews
• Who is an expert?
  - questionnaire design expert
  - subject matter expert
  - questionnaire administration expert (e.g. interviewers)
  - computer-based expert systems (QUAID)
    - http://quaid.cohmetrix.com/
• What do experts do?
  - identify potential response problems
  - make recommendations for possible improvements
• How do experts do it?
  - individual setting or group setting
  - open-ended comments or problem codes based on formal appraisal system
• What in the end product?
  - individual report of open-ended comments and recommended revisions for each individual question
  - individual report of problem codes to each individual question
  - A summary report of noted problems and revisions
  - A summary report of distribution of problem codes
  - qualitative in nature!
• Expert reviews good at
  - identify problems related to data analysis and question comprehension
  - considerable agreement with latent class analysis results
  - considerable agreement with true records
    - items identified as problematic did have high item non-response and higher level of inaccuracy report

- relatively cheap (and faster)
• Practice Varies
- not much literature
- see table 1 in yan (2012)
- large inconsistencies / disagreement between experts

Focus Groups
• Small Group (5-10) people brought together to
- investigate topics/research questions
- what they think
- examine reality about which people will be asked
- uncover assumptions about reality
- evaluation survey questions
- how they think
- vocab, terms used to think about key concepts
- explore alternative wordings/terms
• Interviewer and respondents don't always agree
- sources of income in low SES neighborhood
- investigator: job salary, interest income, dividends…
- respondents: ad hoc work, illicit activities (drugs, prostitution, gambling)….
• Recruitment: Participants like you target population
- homogeneity v.s. heterogeneity
- each group should be diversified but not too different
- each group have a mixture
- groups could be different with regard to one key characteristic (stratifying sample)
• Moderator's guide
- Translate project purpose into issues
- determine the best flow for the guide
- decide on the type of guide
- write good questions
- open-ends!
- revise and finalize (test out questions, determine length, prioritize)
- keep in mind
- what problem is being solved
- what information is being sought
- what purpose is being served
- success of focus groups depends on a good moderator's guide
- typical flow
- ground rules: what to expect? / confidentiality issues / informing video (audio)-taping
- introductions: help develop rapport
- opening questions: easy low anxiety
- in-depth investigations: deep probing, spontaneous comments
- closure: clarify/validate positions, leave them with good impression
- Good questions guides:
- ask short questions to get long answers
- use words participants would use when talking about issue
- ask questions that are easy to say
- address one issue at a time
- use a conversational tone
- ask open-ended questions
- ask uncued questions before cued questions
- ask positive before negative
- avoid leading questions
- avoid asking "why" (ask "what", "how" instead): unless "why" is what you want to learn
• What focus groups can do
- provide qualitative information

- provide a range of information (variety)
• what focus groups can not do:
- generalize to a large population
• Advantages: efficient, small groups
• Disadvantages
- costly: recruitment, incentive, renting of facility, paying professional moderator, transcribing
- group dynamics may not be desirable
• Lots of textbooks
- Kruger & Casey, 2000; Morgan 1996

Pretesting Techniques: Cognitive Interviews

Cognitive Interviewing
• "….the administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that is author intends."
• "Cognitive Interviewing techniques are used to study the manner in which targeted audiences understand, mentally process, and response to the materials we present – with a special emphasis on potential breakdowns in this process" (Wilis, 2005, p3)
• "The cognitive interview is usually conducted within a broader testing environment, as part of a sequence of activities that, in aggregate, are referred to as cognitive testing" (Ibid, p. 5)
• Focuses on the cognitive processes that respondents use to answer a survey question
• Model of survey response process as a guiding framework
• Small convenience sample representing some of the characteristics of the survey population
• Interviewers who are specially trained to conduct this kind of interview
• often conducted in a lab environment, not out in the field

Cognitive Interviews – Techniques
• Think Aloud
• Verbal Probing (Concurrent/retrospective)
• Vignettes, card sorts, ratings tasks
• Paraphrasing
• Response Latencies

Think-Aloud
• Based on verbal report methods (Ericsson & Simon)
• In responding to a question, the respondent is asked both to answer the question and to verbalize his/her thoughts. The verbal think-aloud stream is analyzed to determine the processes the respondent used to construct an answer to the question
• Assumes…
- reports on thought process available to R
- Rs can report these thought processes
- reporting process does not change the thought process
• Example question presented to respondents to prompt them to think aloud about their answer:
• Make the participant comfortable; acknowledge that "thinking aloud" is an unusual behavior
- societal norms actually train us to avoid this type of behavior!
• Reassure the participant that there are no right or wrong answers
- it is okay to admit that he/she used heuristics to generate a response
• An example typically used to train respondents to think-aloud
- "Try to visualize the place where you live, and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about." (by Mingay)
• Example
- Question: How many times have you talked to doctor in the last 12 months?
- Response: "I guess that depends on what you mean when you say 'talked"'. I talked to my neighbor, who is a doctor…I go to my doctor about once a year for a general check-up… I've also probably been to some type of specialist [bad knee, coughing]… which I'm pretty sure was in the past year…I've also talked to doctors…when I

brought my kids in to the pediatrician…assume you don't want that included…saw a chiropractor…I don't know if that's a doctor…"
  - what did we learn?
    - what R included in answer
    - questions about what counts or not
    - uncertainty about past year
• Outcome
  - informal qualitative assessment
  - coding schemes (example)
    - R changes question to fit knowledge
    - recall episodes about self
    - positive qualitative frequency assessment about self
    - statement about what self normally does
    - trait knowledge about self
  - automated coding
    - audio-recording
    - transcription and code specification
    - computer-based coding
• Pros and Cons
  - Pros
    - standardized probe
    - simple technique to implement
    - interviewer effect on the data are minimized
    - assumed to reflect thought process
  - cons
    - difficult, burdensome for respondents
    - Without additional probing, you may get little information that is useful or a lot of information that is irrelevant
    - cognitive burden of thinking aloud could change how respondents are processing the survey questions

Concurrent/retrospective Verbal Probing
• Concurrent Probing
  - interviewer asks the survey question
  - R answers the question
  - Interviewer asks a probe question
  - R answers the probe question
• Retrospective probing
  - interviewer asks (and R answers) all the survey questions
  - interviewer asks (and R answers) probe questions
• Examples
  - Comprehension/Interpretation: What does the term outpatient mean to you?
  - Paraphrasing: can you repeat the question I just asked to you?
  - confidence judgement: How sure are you that your health insurance covers drug treatment?
  - recall probe: how do you remember that you went to the doctor five times in the last 12 month?
  - specific probe: why do you think cancer is the most serious health problem?
  - general probe: how did you arrive at that answer? Was it hard to answer?
  - willis (2005): good resource for probing examples
• Proactive probes
  - researcher actively probing to search for problems
  - what to probe may be specified in advance of interview
  - scripted in advance (anticipated) or left to interviewer to formulate (spontaneous)
  - Examples
    - Remember to tell me what you are thinking
    - what makes you think that?
    - what were you thinking about as you were trying to remember that?
    - what types of events did you include in your answer?
• Reactive probes

- based on R answers that suggest a problem
- standard probes from a prepared set can be presented as needed
- probes can be scripted in advance (conditional) or on the spot (emergent)
- Examples
  - you look puzzled, what are you thinking?
  - you answered that very quickly. Why was it that easy for you to come up with your answer?
  - I appreciate the effort you're making to recall this information

Cognitive Interviewing Good At….
• Uncovering problems with
  - comprehension
  - difficulty with recalling and estimation/judgement
  - willingness to carry out recall, estimation/judgement task
  - qualitative in nature
• But not so good at
  - interviewer problems
  - analytic problems
  - generalizing to a large population
• Also… costly and personnel issues

Cognitive Interview: Last Words
• Practice varies by organizations
• results vary by cognitive interviewers, administrations, organizations
• Lots of good textbooks: Willis (2005)

Behavior Coding

Behavior Coding
• Methodical observations of Interviewer and R behaviors from audio recorded interviews
• Provides objective measures of problems with Qs
• Systematic, replicable, reliable
• Code observes I-R interaction: Live or recorded interviews, in-person or telephone
• Assigns a code to, i.e. classifies observed problem
• Coding over multiple interviews by multiple interviewers
• Can help understand the question-answer process
  - deviations from ideal Q-A sequence ➔ problem
• Quantitative summary of codes to identify problematic questions

When is behavior coding used and for what purpose?
• Prior to actual data collection:
  - pre-testing questions
  - pre-testing data-collection procedures
• During data-collection
  - monitoring interviewers
• After data collection
  - evaluating data quality
  - explore causes and effects of behaviors

Paradigmatic(ideal) Q-A Sequence (Maynard & Schaeffer, 2002)
• I: How many days a week do you exercise?
• R: Seven days
• Okay, Thank you
➔ Behavior coding tries to detect any deviations from this kind of ideal Q-A flow

Non-Paradigmatic Q-A Sequence – Example #1
• I: How many days a week do you exercise?

• R: Excuse me?
• I: How many days a week do you exercise?
• R: Seven days
• I: Okay, thank you

Non-Paradigmatic Q-A Sequence – Example #2
• I: How many days a week do you exercise?
• R: Hmm, most days?
• I: Six days a week?
• R: Yes
• I: Okay, thank you
➔ Interviewer's "directive probe" (answer not from the Respondent) ➔ be surely flagged by most behavior coding schemes

Examples of Behavior Codes for Interviewer and Respondent Behaviors

Code Category	Description
Interviewer Questioning Behaviors (choose one)	
1.	Reads question exactly as worded
2.	Reads questions with minor changes
3.	Reads questions so that meaning is altered
Respondent Behaviors (check as many as apply)	
1.	Interrupts question reading
2.	Asks for clarification of question
3.	Gives adequate answer
4.	Gives answer qualified about accuracy
5.	Gives answer inadequate for question
6.	Answers "don't know"
7.	Refuses to answer

Source: Groves et al. (2009)

Oksenberg, Cannell & Kalton (1991)

1. What was the purpose of that visit (to a health care person or organization)?
2. How much did you pay or will you have to pay out-of-pocket for your most recent visit? Do not include what insurance has paid for or will pay for. If you don't know the exact amount, please give me your best estimate.
3. When was the last time you had a general physical examination or checkup?

Percent of Problems per Question

Question	1	2	3
Interviewer action			
Slight wording change	2	30	3
Major wording change	3	17	2
Respondent action			
interruption	0	23	0
clarification request	2	10	3
inadequate answer	5	17	87
"don't know"	0	8	12

When behavior coding is used for pretesting…
• Respondent codes that are most useful
  - request for clarification/request to repeat
  - qualified answer, inadequate answer, don't know/refusal
  - problems with questions may be visible in very subtle ways
• Interviewer codes that are most useful
  - minor/major changes in question wording
  - might be more important to know what is changed in question wording
• Benefits of behavior coding: reliable and quantitative

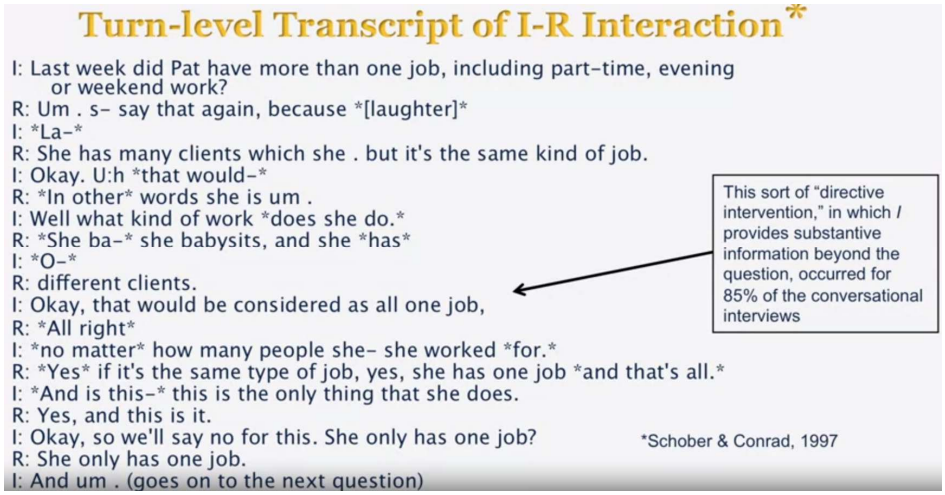
Behavior Coding – Unit of Coding
• Roughest Unit: whole interview
• Most frequently used: QA sequence
• Intermediate level: Exchange
• Most detailed level: Utterance or Turns
  - selective coding: only utterances that are within the set of pre-specified codes are coded
  - full coding: all utterances are coded

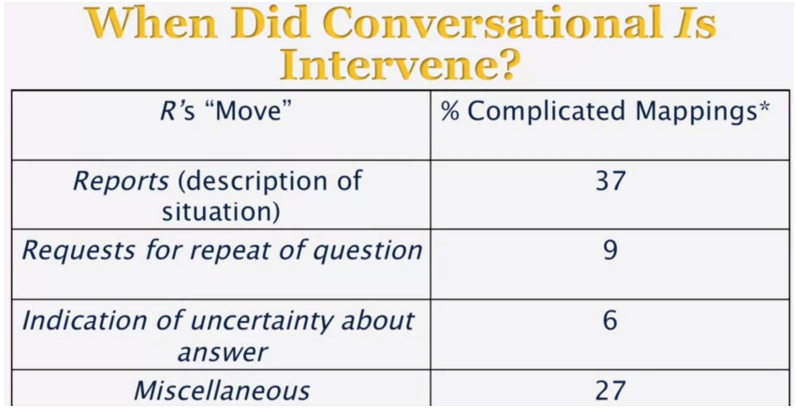
Behavior Coding – Unit of Analysis
• Question Level
  - Q5: x% R requested for clarification
  - Q7: y% of Interviewer had minor wording changes
• Respondent Level
  - Older Rs requested clarification more frequently than younger respondents
• Interviewer level
  - Interviewer A had major wording changes for 10 questions

Turn Level Analysis
• In what sequence of conversational turns did the problem appear?

- did / re-read the question when R's answer not one of the response categories?
- Did R answer the question after receiving a neutral probe, e.g. "Let me re-read the question?"

• Can help explain how interviewing techniques are working or not





n = 64

Behavior Coding shortcomings

• Behavior coding reveals only observable problems
• Behavior coding may not provide information on underlying causes of problems
  - does not explain source of the problems and so results may not extend to other questionnaires, help develop general principles etc
• May miss important interactive phenomena (at QA sequence level)
  - hard to know what behavior by / preceded (and possibly contributed to) R's behavior
• Can be labor intensive and not optimal for production deadline situations
• Won't tell designers how to fix problems

Behavior Coding: Summary

• Interviewer-oriented codes
  - reading error
  - other probing behaviors
• respondent-oriented codes
  - R request for repeat
  - R requests clarification
  - R provides uncodeable response
  - R interrupts
  - R is uncertain
• Behavior coding tells us **how frequently** problems occurred
• But **doesn't** directly indicate **why** they occurred, or the exact nature of the problem

- If interrupted, where did this happen and why?
- If respondents sought clarification, what did they ask about?

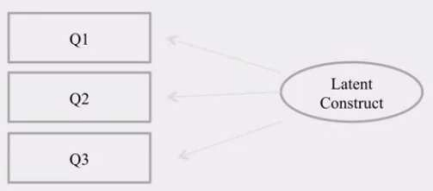
• Fowler: Use (qualitative) coder debriefing
- interview coders after having listened to and coded numerous interviews
- they tend to know what is happening

5.2. Quantitative Pretesting Techniques

Pretesting Techniques: Quant Techniques

Latent Class Analysis (LCA)

• Multiple indicators (aka. Survey questions) for a (latent) construct



• Indicators do not have to be error free

• However, errors associated with indicators are assumed independent conditional on latent variable ('local independence' assumption)

• Estimating unconditional and conditional probabilities

• Two conditional probabilities are important for question testing and evaluation
- false positive
- false negative

• Indicators with higher error rates are assumed to be 'bad'

Indicator	Latent Class	
	c=1	**c=2**
u_1=1	$P(u_1 = 1\|c = 1)$	**$P(u_1 = 1\|c = 2)$**
u_1=2	**$P(u_1 = 2\|c = 1)$**	$P(u_1 = 2\|c = 2)$
Unconditional Probabilities	**P(c=1)**	**P(c=2)**

• Data collection needed: sample size can't be too small

• For a latent class with 2 classes, three items are needed in order for model to be identified

• Assumptions have to be made to achieve identifiability
- biased estimates of error rates when assumptions are not met

• can identify 'bad' survey item

• but doesn't indicate why and doesn't suggest fix

SQP

• Built on MTMM experiments conducted on multiple items, across multiple countries
- generated validity, reliability

• coded question characteristics

• fit regression models

• when you use the program,
- you enter your survey questions
- SQP
  - code your questions
  - produce (predicted) validity and reliability

Field Test
• "conventional pretest", "dressed rehearsal", "pilot study"
• Implement the questionnaire on a small sample
  - 15-35 respondents similar to your actual respondents
  - mostly convenience sample
• adopt similar data collection protocol
• goals are to find out
  - "practical problems"
    - interviewers? Respondents?
    - length: question level, section level, full instrument
    - examining key variables: distribution, tabulations, missing data

Field test – Add Ons
• With field test
  - building in behavior coding
  - building in cognitive probes
  - building in respondent debriefing
  - building in interviewer debriefing
  - starting statistical analyses:
    - validity
    - reliability
    - latent class models
    - structural equation models
• Costly – especially with add-on

Pretesting Techniques: Other techniques

Other Pretesting/Testing Methods
• Interviewer debriefing
  - individual setting: Interviewer questionnaire (Fowler 1995)
  - group setting
  - notes/comments interviewer noted on the questionnaire during administration
• Interviewers good at uncovering:
  - skipping errors, typos, misalignment
  - Qs difficult for Rs or Rs unwilling to answer
  - length of questionnaire/particular questions
• Marginal cost to hold interviewer debriefing!
• Respondent debriefing
  - embed additional questions in questionnaire
    - general
      - which questions were difficult to answer and why?
      - which questions were too personal?
      - (SAQ): is the form too confusing?
    - specific probes on specific questions
      - similar to probes used in cognitive interviews
    - do not turn it into a cognitive interview!
  - In lab setting, PI/survey methodologist talk to R or Rs (very rare)
• Question wording experiments
  - split ballot (schuman and presser 1981)
  - randomization of respondents to one of conditions
    - "Do you think the U.S. should forbid public speeches against democracy?" v.s.
    - "Do you think the U.S. should allow public speeches against democracy?"
    - comparing response distributions
• Vignettes
  - hypothetical situations
  - factors/dimensions can be varied

Comparing Pretesting Methods
• Three general approaches to evaluation (R. Caspar)
  - explanatory
    - compares methods in terms of effectiveness for detecting problems
  - confirmatory
    - compares methods in terms of effectiveness for confirming / disconfirming problems that are suspected based on other results
    - reparatory
      - compares methods in terms of effectiveness for suggesting revisions that will improve the data

Summary: Expert Reviews
• Pros
  - cost effective
  - identifies a wide variety of potential problems from typos and skip pattern logic errors to problems in how concepts have been operationalized
• Cons
  - reviewer dependent
  - no input from actual respondents

Summary: Focus Groups
• Pros
  - useful when there is little information on the topic or on a particular subgroup of subjects
  - the approach is flexible enough that unexpected information can be immediately followed up on by the moderator
  - information is obtained directly from the types of individuals who will participate in the study eventually
• Cons
  - small group dynamics must be appropriately controlled or results have limited value
  - analysis can be time consuming
    - qualitative review of the transcript
    - quantitative analysis using qualitative software
    - results from a small number of subjects must be interpreted with care

Summary: Cognitive Interviews Think-Alouds
• Pros
  - freedom from interviewer imposed bias
    - no additional comments from interviewer
  - minimal interviewer training
    - merely read instruction, no further intervention
  - open ended format: allows for unanticipated comments
• Cons
  - need to train respondents
  - ability to participate in think-aloud interview correlated with education
  - burden on subject
  - easy to diverge from task
  - answers not always useful
    - unclear if respondent would be able to answer
    - success question dependent (no strategizing needed)
    - potential violation of non-reactivity assumption

Summary: Cognitive Interviews Probing
• Pros
  - comments come from respondent directly
  - question-specific comments can be used
  - to rectify or eliminate problematic questions
  - interviewer can focus on particular problems
  - interviewer controls the discussion
  - easy to introduce to respondents

• Cons
  - potential for reactivity with in-between probes
    - interview flow might get disrupted (alter answers)
    - lengthen the interview
  - debriefing at the end of the questionnaire less useful
    - respondent may only want to end the interview by this point so lengthy commentary is unlikely
  - underlying assumption: Rs are willing to admit confusion

Summary: Latent Class Analysis
• Pros
  - Qualitative conclusions on different questions
  - Quantitative estimates of error rates
  - No gold standards are needed
• Cons
  - estimation of error rates dependent on assumptions
  - violation of assumptions not known or testable

Summary: Interviewer Debriefing
• Pros
  - includes interviewer evaluation of what makes a question or procedures difficult to implement
  - takes advantage of the varied interviewing situations and types of respondents the interviewer experiences
• Cons
  - interviewer's attribution of respondent difficulty may be incorrect

Summary: Vignettes
• Pros
  - can be a less threatening way to study sensitive topics
  - systematic manipulation of key variables in the vignette allows for quantitative analysis
  - iterative approach allows researcher to hone definitions or descriptions to maximize consistent understanding
• Cons
  - responses are not based on actual subject behavior or opinion
  - artificial nature of the vignette may not accurately reflect "real world" behaviors
    - there may be a disconnect between what the subject says he/she would do and actual behavior

Write an organized testing report
• intro section with purpose of pretest
• nature of the subject population and recruitment process
• Number of subjects tested with tabular summary
• Describe number and experience of interviewers
• Techniques used (e.g. for cognitive interviewers, probing categories such as concurrent, retrospective, proactive v.s. reactive, degree of probe standardization etc.)
• Q by Q review of results, with recommendations
• Section of limitations of the testing, next steps