Class 2: Principles of Questionnaire Design

MA5953: Creating Your Own Data

Dr. Miriam Sorace

www.miriamsorace.github.io

6 November 2023

Quick Recap: Last Class

- Course structure and website
- Different types of surveys
  - Structured
  - Semi-structured
  - Unstructured
- How to write a report using RMarkdown
  - Downloading files from GitHub
  - Knitting the RMarkdown file
  - Do check out the RMarkdown Cheat Sheet and practice with it on your own!

Recap: The Importance of Survey Research

The COVID-19 Social Study

Outline of Today's Class

Validity & Reliability

Formulating Questions

- Consider the following survey item examples, aimed at measuring social media use and detecting social media over-use:
  1. *"How many hours did you spend on Twitter yesterday?"*
  2. *"How many hours per day have you spent - on average - on social media sites in the past year?"*
- Which alternative would you pick? Would either one of them work? If not, why do you think the survey item is not appropriate to measure the level of social media use?

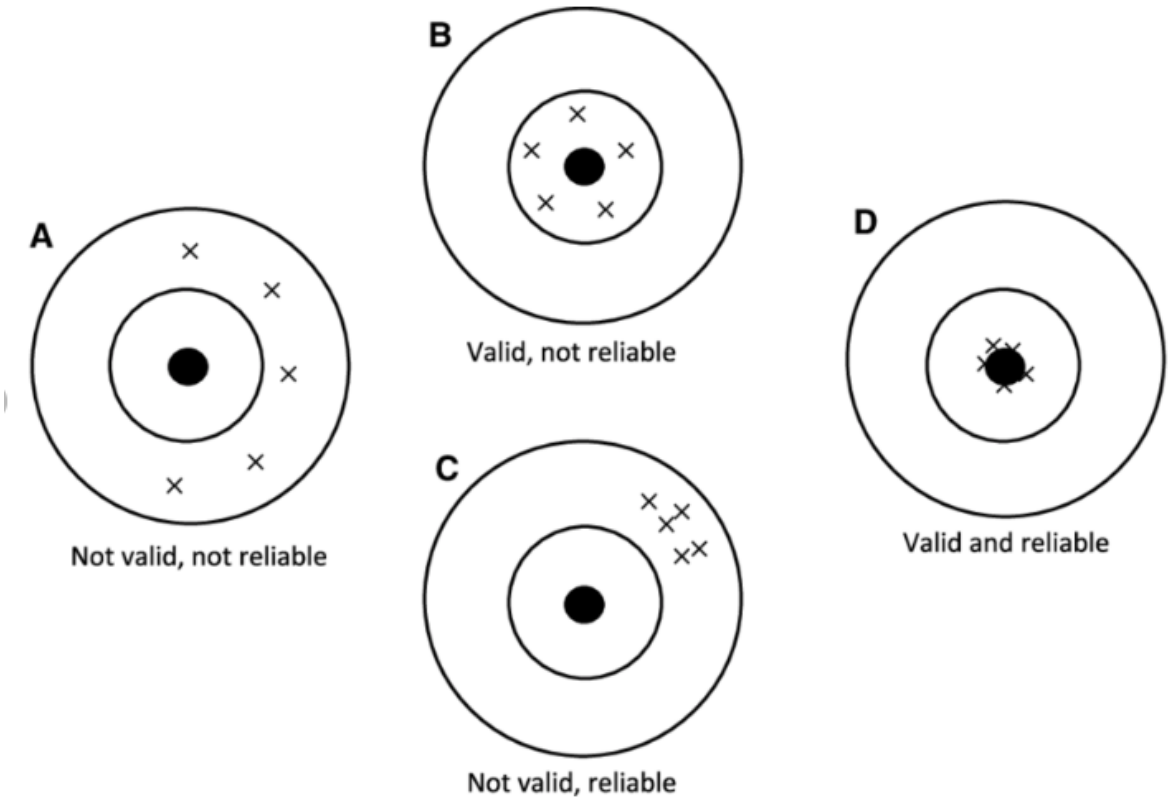
Validity

- The measure (i.e. the survey item) is related to the underlying construct.
  - E.g. the response delivers the true characteristic/behaviour/attitude that the researcher intended to elicit.

Reliability

- The measure (i.e. the survey item) performs consistently
  - E.g. the response to the question is consistent within the same individual over-time (*test-retest*), or different individuals respond using the same response process (*test-test*).

Validity vs. Reliability



This figure shows possible combinations of validity and reliability. Each 'target' contains a bullseye, which represents the real value of the measured property. A measure is reliable if repeated measures (×) fall within the same small area on the target. A measure is valid if repeated measures are clustered around the bulls-eye. A measure is more accurate the closer the x's are to the bulls-eye. a Shows an unreliable and also invalid measure. b Shows an unreliable but valid measure. c Shows a reliable but invalid measure. d Shows a measure that is both valid and reliable. Unreliable measures are either not valid (a) or inaccurate (b). However, reliable measures are not always valid (C)

1

[1] Image taken from: Irvine, Elizabeth. (2019). Developing Dark Pessimism Towards the Justificatory Role of Introspective Reports. *Erkenntnis*

Assessing Validity & Reliability

Going back to the examples

- Underlying construct = social media (over)use:

  1. *"How many hours did you spend on Twitter yesterday?"*
  2. *"How many hours per day have you spent - on average - on social media sites in the past year?"*

Example 1: Construct Validity

- 'Twitter' $\neq$ *all* social media
- 'Yesterday' helps recall and precision, but might not be representative of general consumption
- This might be a case where reliability works well, but the capacity of the item to measure the underlying concept (social media use) is low (e.g. *low construct validity*). Like scenario C in the figure above.

Example 2: Reliability

- 'the past year' might be too long a time-frame for precise recall.
- calculating averages might have different degrees of complexity for different respondents
- 'social media sites' may be too broad: without a list helping recall, we are not sure that two people will include the same sites in their calculations or that even the same person will use the same sites from one time to the next
- This might be a case where validity might be approached, but the capacity of the item to be consistent is now (e.g. *low reliability*). Like scenario B in the figure above.

Testing Validity

- Correlation between the survey answers and some measure of the *actual* underlying construct
- Two main methods:
  - **Gold Standard**: data external to the survey
    - E.g. health responses vs. actual medical records
  - **Theoretical Correlate**: use responses to attitude/demographics that should be theoretically related
    - E.g. ideology question correlated with party ID self-report

Testing Reliability

Main method:

- Compare results from repeated interviews (better for more stable constructs - e.g. personality)
  - E.g. same people interviewed twice or aggregate results compared from 2 different interviews

The Survey Response Process

The Survey Response Process

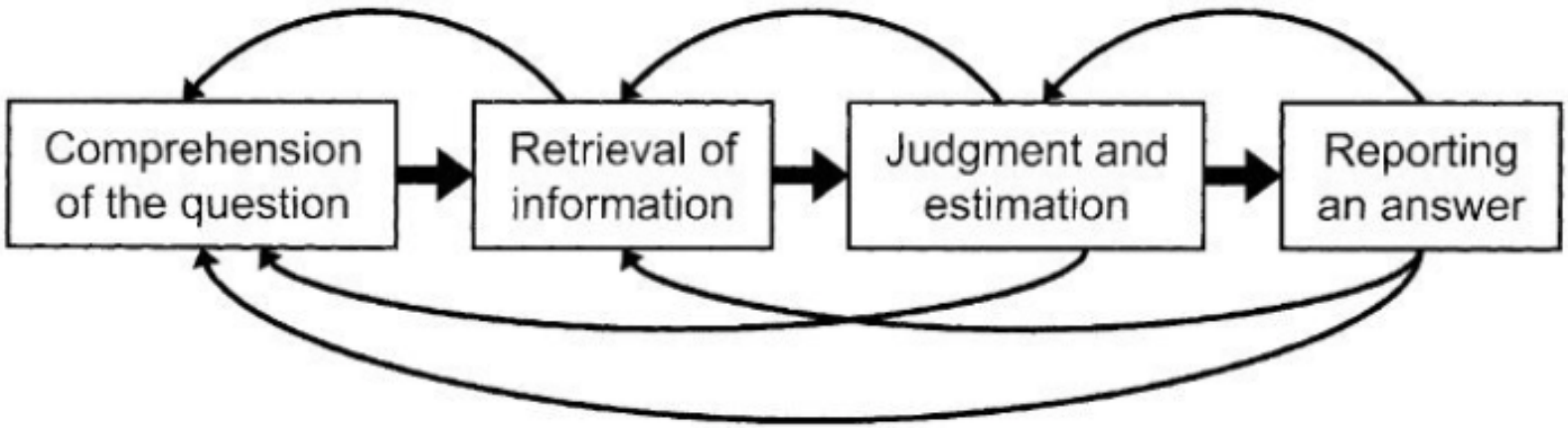


Figure 7.1 A simple model of the survey response process.

The Survey Response Process

- **Comprehension:** interpretation of the survey item
  - Parsing/reading
  - Assigning meaning to component concepts
  - Inferring the purpose 'behind' the question (pragmatics: inferences from co-text and context)
- **Retrieval:** recall of all pieces of information required to answer the question
- **Judgment:** combination/summary of the pieces of information recalled
- **Reporting:** formulation of the response in the required format

The Survey Response Process

Potential Problems in Answering Survey Questions

- **Comprehension:** Misunderstandings can occur
  - Avoid ambiguity:
    - 'You' is not as unambiguous a word! (singular or plural?)
    - 'week-end': does it include Friday?
  - Avoid jargon or acronyms: don't pre-suppose knowledge. If such terminology is necessary, do provide a definition cue.
  - Avoid complicated sentence structure
  - Avoid 'leading' questions (presupposition) or order effects
    - "you must know that X happened. What do you think about Y's response?"
    - "The government encourages citizens do vote. Did you turn out to vote in the last election?"
    - Do not ask about specific things (e.g. happiness in marriage) before general things (e.g. happiness in general) or people will exclude married life from the latter
  - Ask one question at a time: reduce complexity

The Survey Response Process

Potential Problems in Answering Survey Questions

- **Retrieval:** ample evidence that individuals take shortcuts / are 'satisficers' (want to give best answer available at lowest cognitive cost) instead of 'optimizers' (answering comprehensively)
  - Avoid asking questions too far back in time: memory decays over time, use long reference periods only for very memorable events
  - Provide *retrieval cues*: decompose various types of event instead of bunching them up in a general category, offer participant cues or location cues to aide memory
  - Use personal landmarks or life event calendars; use distinctive/important/emotionally involving events

The Survey Response Process

Potential Problems in Answering Survey Questions

- **Judgment:** respondents may be led astray in the process of combining items of information and deciding on their relevance.
  - Avoid double-barreled questions (i.e. questions that link together two attitude objects). E.g. from Schuman & Presser (1981):
    - "Do you agree with the US government decision to defend Korea?"
    - vs. "Do you agree with the US government decision to defend Korea to stop a Communist take-over?"
    - the second wording made Communism salient and heightened support for the Korean War. Different considerations clearly entered the equation.
  - Avoid sensitive questions without adding 'forgiving wording' as a preface, e.g. "we often find a lot of people that ..." (*'everybody does it' approach*) or by providing reasons for undesirable behaviours. Using open-ended questions for those can also help.
    - *Social Desirability Bias*:The under-reporting of socially undesirable behaviours and over-reporting of socially desirable ones.

The Survey Response Process

Potential Problems in Answering Survey Questions

Reporting: Individuals are satisficers, so the response options have to be considered carefully, and potentially randomised in their order.

- Randomise question or response order.
  - *Response Order Effects*: picking the most available answer choice and/or answering a subsequent question in reference to preceding questions.
- Show labels in sliders + avoid negative numbers in scales
  - *Positivity bias*: unwillingness to pick negative ends of the scale, or extreme answer categories
- Avoid truncation
- Do not have too many answer categories: 5 or 7 point scale often the best compromise.
- Avoid ranking tasks: they often exceed cognitive capacity of respondents, paired comparisons might be better/easier.
- Avoid creating "Check all that apply" items, ask respondents to answer yes or no for each.

Key Terms/Concepts

- **Double-barreled question:** questions that link together two attitude objects). E.g. from Schuman & Presser (1981):
  - "Do you agree with the US government decision to defend Korea?"
  - vs. "Do you agree with the US government decision to defend Korea to stop a Communist take-over?"
- **Satisficing:** producing best answer available at lowest cognitive cost vs. 'optimizing' (answering comprehensively)
- **Social Desirability Bias:** The under-reporting of socially undesirable behaviours and over-reporting of socially desirable ones.
- **Positivity & Acquiescence Biases:** unwillingness to pick negative ends of the scale or disagree options, or extreme answer categories

Best Practices in Question Design

Pilot, Pilot, and Pilot again!

- "Survey researchers evaluate question comprehension, difficulty in memory retrieval, and related issues primarily by observing people trying to understand and answer the questions." (Groves et al. 2009: 259)

Key Approaches

- Expert Reviews
- Focus Group Discussions
  - Qualitative interviews of target population in groups
- **Cognitive Interviews**
  - Individual qualitative interviewing
- Field pre-tests
  - small N of interviews (typically 100) using sampling and method intended for the full-scale survey and including de-briefs with participants. Behaviours (frequency of DKs, item non-response, answer distributions) used to test.
- Randomized 'split ballot' experiments
  - mostly used to pre-test different wordings of the same question

Cognitive Interviews

The standard pre-test

- Protocol analysis, introduced by Ericsson and Simon (1980)
- Interviewees 'think aloud' - or answer probing questions
- Alternatives to verbalisation:
  - Confidence ratings: respondents asked to assess confidence in their answer
  - Paraphrasing: respondents asked to re-formulate the question in their own words

Cognitive Interviews

Process

- Build useful **probes**
  - Questions/statements that encourage the respond to extend/elaborate on an answer. E.g. (1) Non-directive probing ("tell me more about it . . . "; "what do you mean by that?") OR (2) Summary technique (mostly in semi-structured and unstructured interviews)
- Those that follow Tourangeau's (1984) Survey Response Model are best:
  - "what does [concept x] mean to you?" (*Comprehension*)
  - "what information/knowledge did you collect/search in your memory to answer this question?" / "How did you decide which answer to give?" (*Retrieval*)
  - "how hard was this to answer?" / "was this question hard or easy to answer? Why?" (*Judgment*)
  - "Were the response categories adequate/sufficient?" (*Response*)

Cognitive Interviews

Reporting Results

- Frequency of 'don't know' mentions
- Average time taken to answer the question
- Comprehension probes: thematic analysis, e.g. hand-coding on the basis of a codebook
- Recall probes: count number of steps mentioned by respondent.
- Difficulty probes: frequency saying "hard".

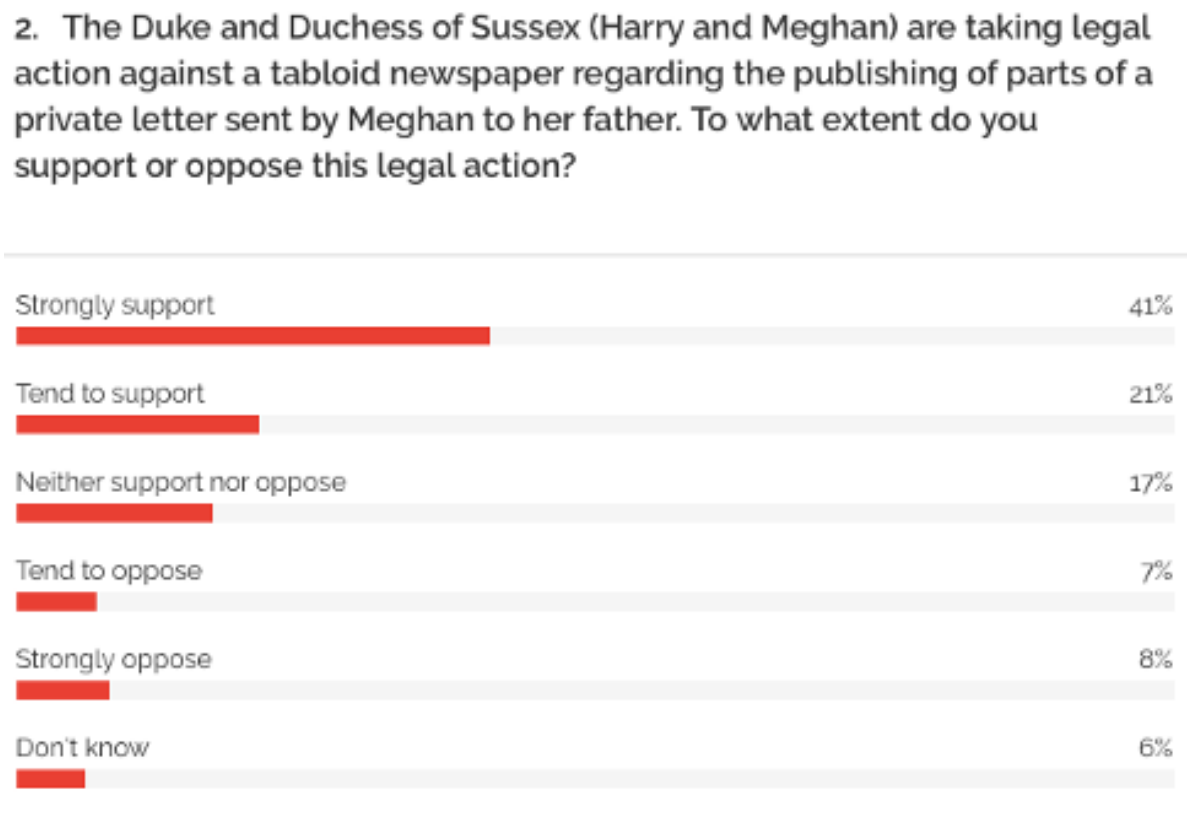
Cognitive Interviews

Best Practice. From: WDI UMich: Cognitive Interview Guide

- Inform the participant about confidentiality and anonymity procedures
- Respect participants' time: do not keep participants past the agreed time.
- Do not force participants to answer, and reassure that there is no right/wrong answer.
- Clarifications/question repetition might be requested. Note how often this happens, valuable data!
- Always thank the participant for their time!

Lab Work: Evaluating Survey Items

Applied Task 1: Survey Item Evaluation



[2]

[2] Taken from: https://yougov.co.uk/opi/surveys/results#/survey/1144947a-e4f8-11e9-ad60-47ee57021fb0

Applied Task 1: Survey Item Evaluation

QA3	How important for you personally are the following points?				
Public officials and politicians do not use their positions to obtain benefits for themselves or their family members but take decisions in the public interest	1	2	3	4	5

[3]

[3] Taken from: https://search.gesis.org/research_data/ZA7572?doi=10.4232/1.13372

Applied Task 1: Survey Item Evaluation

How many children do you have?

- 1-2
- 3-4
- 5-6
- 6+

Applied Task 1: Survey Item Evaluation

QA3	How important for you personally are the following points?					
	(READ OUT - ONE ANSWER PER LINE)					
		Essential	Important	Not so important	Not important at all	DK

[4]

[4] Taken from: https://search.gesis.org/research_data/ZA7572?doi=10.4232/1.13372

Applied Task 2: Formulating a Question

10 mins

- In groups of 3 or 4: design a survey question on **one** of the below constructs:
  1. Attitudes on Israel-Palestine war
  2. Happiness levels
  3. Recycling behaviour
  4. Evaluation of personal economic situation
  5. Attitudes on gender equality
  6. Attitudes towards supporters of one's opposite political party
  7. Healthy eating behaviour

Applied Task 3: Cognitive Interviewing

5 mins

- Build your cognitive interviewing protocol:
- Pick 4 cognitive probes (1 for each step of the survey response process), following the example here:
  - From WDI UMich: Cognitive Interview Guide

Applied Task 3: Cognitive Interviewing

10 mins

- Elect a representative from your group that is going to deliver the cognitive interview
- Exchange your original survey item with the pair/group next to you
- Evaluate each other's original survey items: read the survey item and respond to the interviewer's questions!

Applied Task 3: Evaluation

5 mins

- What did you learn from this exercise?
- Did you get ideas on how to reformulate the original survey question?
- What are the key step of the survey design process?

For Next Time ...

- Refine your survey item: what do you want to capture? How can you approach validity?
- Design your own cognitive interview: what probing questions would you field?

Borrowing from Gold Standard Surveys

Don't re-invent the wheel!

- You can check survey items previously run by other surveys here:
  `https://www.icpsr.umich.edu/web/pages/ICPSR/ssvd/`
- Or you can play around with the Survey Quality Predictor
  `http://sqp.upf.edu/`

What did we learn today?

- Validity & Reliability
- The Survey Response Process
  - Comprehension
  - Retrieval
  - Judgment
  - Reporting
- Biases in answering survey questions
- How to test for validity via cognitive interviews