

# Questionnaire design course

Day 2



# Day 2

- Evaluating survey questions
- Answer formats and rating scales
- Writing questions for different modes

# DAY 2

# Evaluating survey questions



# Evaluation methods

Quantitative	Desk
Standard field testing (pilot)	Expert review
<b>Respondent debriefing/ web probing</b>	<b>Systematic forms review</b>
Vignettes	<b>Qualitative</b>
Behaviour coding	Focus groups/depth interviews
Split ballot experiments	Card sorts/vignettes
Analysis methods (e.g. latent class analysis, Item Response Theory, Multi-Trait-Multi-Method experiments)	<b>Cognitive interviewing</b>
Eye tracking	Usability testing
<b>Can also combine methods</b>	

# Desk based appraisal

- systematic review of questionnaire
- focuses on both respondent & interviewer anticipated problems
- codes used to identify problems
  - grounded in theory
  - examples: reading, instructions, clarity, assumptions, knowledge and recall, sensitivity/bias, response categories
- Questionnaire Appraisal System (QAS 99) is most widely used

Copy & paste this link into your web browser

<https://www.cdc.gov/healthyyouth/evaluation/pdf/brief15.pdf>

# Extract from QAS 99

<b>STEP 3 - CLARITY: Identify problems related to communicating the <i>intent</i> or <i>meaning</i> of the question to the respondent.</b>		
3a. <b>WORDING:</b> Question is lengthy, awkward, ungrammatical, or contains complicated syntax.	YES	NO
3b. <b>TECHNICAL TERM(S)</b> are undefined, unclear, or complex.	YES	NO
3c. <b>VAGUE:</b> There are multiple ways to interpret the question or to decide what is to be included or excluded	YES	NO
3d. <b>REFERENCE PERIODS</b> are missing, not well specified, or in conflict.	YES	NO

# Practical exercise [15 mins]

Use the cut-down QAS to review question

- What did the review highlight?
- How did you find using the QAS?

# Respondent debriefing

---

- Follow up questions designed to determine respondents' understanding of the survey question
- Focus on comprehension, recall, judgement, response
- Questions can be open or closed



# Example follow up question

Follow up to a series of Qs about main job, where respondent indicated had more than one job.

How did you decide which job was your main job?

Job worked at the most hours

The one that pays the most

The one been in the longest

Other (please specify)

# Practical exercise [5 mins]

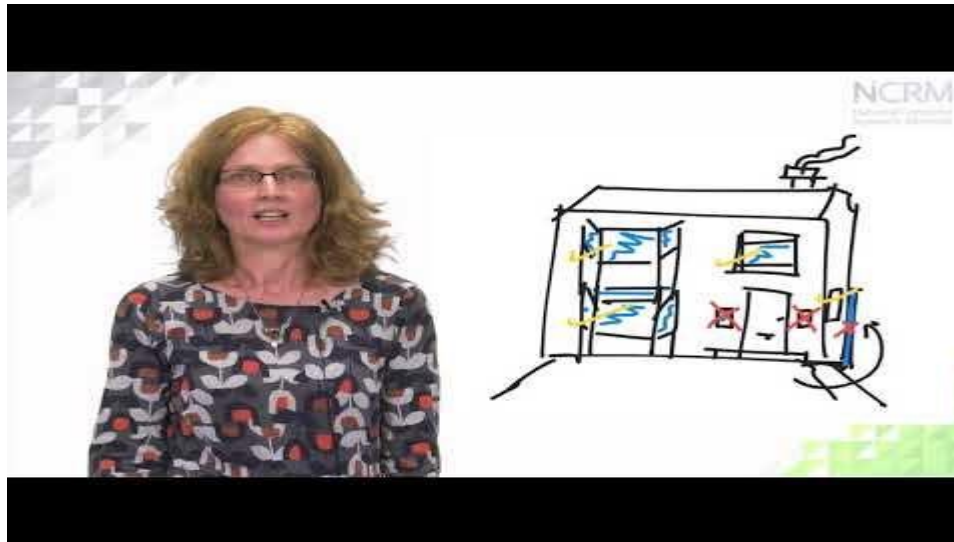
Use your sleep question

- Write a respondent debriefing question for it
- Review and revise
- Share with the whole group

# Pros and cons

Pros	Cons
Large sample size/ sample coverage compared to CI	Writing focused Qs that yield useful data can be difficult
Quantitative results	Increases questionnaire length & respondent burden
	Limited coverage of questions & exploration of problems
(For web probing)	
No interviewer effect	Limited to pre-scripted questions: no opportunities to follow up
	~20% of responses uninterpretable

# Cognitive interviewing – watch this video in your own time



Video can be viewed by pasting this link into your web browser:

<https://youtu.be/OVRgH0OrQPM>

# Example

---

## Survey Question:

- In the last 30 days, did you or other adults in your household ever cut the size of the meals or skip meals because there wasn't enough money for food?

## Measurement objective

- Identify people at extreme risk of food poverty

## Issues to investigate

- Does the respondent know the answer? If so, how sure are they? What did they understand by the term 'meals'? Who did they include as adults in their household?

## Suggested probes

- How did you go about answering this question? Explore anything that made the question difficult to answer
- The question asked about meals. What did you think of as being meals? Ask for examples
- The question asked about you or other adults in your household. Who did you include in your answer? Check age of people included

# Reflections on question testing methods

# Answer formats & rating scales

6.

# Keep in mind – levels of measurement

Nominal	Can you speak Welsh? 1 Yes 2 No
Ordinal	Which of these best describes your ability in spoken Welsh 1 I am fluent in Welsh 2 I can speak a fair amount of Welsh 3 I can only say a few words in Welsh 4 I can say just a few words in Welsh
Interval	On a scale of 1 to 10, where 0 is not at all and 10 is native fluency, how would you rate your Welsh language skills?
	How long was your journey? Hours : minutes <input type="text"/> : <input type="text"/> 00



# Open vs closed questions

Which sport or sports do you participate in? OPEN

Which, if any, of the following sports do you participate in?

List, Other, None of these

# Open vs closed questions

	Pros	Cons
Open	Respondents can answer in their own words	Responses have to be coded to allow statistical analysis, which is costly, time-consuming and subject to error
	Are not leading so can, potentially allow measurement of salience – how important issue is to respondent; indicate respondent's level of knowledge; indicate strength of feeling	If an interviewer-administered survey, can be difficult to get interviewers to probe consistently
	Can avoid format effects, such as primacy (the tendency to endorse the first option seen) or recency (the tendency to endorse the last option heard)	Respondents may not provide sufficient detail when answering to capture key differences
	Are required for the development of response options for closed questions	Can generate irrelevant responses
Closed	Can help with respondent-recall	If not a complete list of answer-options, can introduce bias
	Level of detail, areas of interest can be conveyed to respondent	Cannot capture detailed information, for example occupation details
	Little or no coding of answers required, thus quicker and cheaper	Categories may not be recognisable to respondents
	Less likelihood of coder bias or inconsistent probing	Answer options can influence way in which respondents' interpret question

Closed question	Open question
<p>Which of the things on this list would you most prefer in a job?</p> <ol style="list-style-type: none"> <li>1. High Income (12%)</li> <li>2. No risk of being fired (7%)</li> <li>3. Working hours that are short (3%)</li> <li>4. Chances for advancement (17%)</li> <li>5. The work is important &amp; gives a feeling of accomplishment (59%)</li> </ol>	<p>People look for different things in a job, What would you most look for?</p> <ol style="list-style-type: none"> <li>1. Pay (12%)</li> <li>2. Security (7%)</li> <li>3. Short hours (1%)</li> <li>4. Opportunity for promotion (2%)</li> <li>5. Stimulating work (21%)</li> <li>6. Pleasant/enjoyable work (15%)</li> <li>7. Work conditions (15%)</li> <li>8. Job satisfaction/ liking job (17%)</li> <li>9. Other, DK etc. (10%)</li> </ol>

Source: Schuman, H. & Presser, S (1979) The Open and Closed Question. American Sociological Review. 44 (5) 692-712

# Guiding principles

---

- **Use answer categories when you want respondents to respond in a particular format** i.e. a format that won't necessarily be evoked by the question alone
- **Use open questions when collecting accurate numeric data** (e.g. frequencies, durations or expenditure) or when you want to capture spontaneous, top-of-mind responses
- **Use banded answer categories to collect numeric data when a high level of accuracy isn't required**

# Number of response options

---

- A wealth of literature looks at the 'ideal' number of scale points
- No consensus on the ideal number has been reached (See Preston & Coleman, 2000)
- Most studies put the ideal between 5-11
- Number of points should be determined by measurement objectives
- Can Rs meaningfully differentiate between points?
- Is scale sufficiently sensitive? Rs can experience frustration if scale pushes them to have an 'extreme' view (Cox, 1980)

# Is this rating scale balanced?

How would you rate our service?

1) Excellent

2) Very good

3) Good

4) Fair

5) Poor

1) Very good

2) Good

3) Bad

4) Very bad

# Answers scales for rating questions

---

## ■ Visual analogue scales

Please put a mark on the line to indicate how you feel today?

Very unhappy \_\_\_\_\_ Very happy

## ■ Verbal answer scales

How happy are you feeling today

(Very happy, happy, unhappy, very unhappy)

## ■ End-labelled answer scales

How happy are you feeling on a scale of 1-10.

1 is very unhappy and 10 is very happy?

# Agree/Disagree vs. Construct specific

---

<b>Agree/Disagree</b>	<b>Construct Specific</b>
<p>To what extent do you agree or disagree with the following statement: Overall, I am better off on Universal Credit...</p> <p>Strongly agree, Agree, Neither agree or disagree, Disagree Strongly disagree</p>	<p>Since you started claiming Universal Credit would you say your financial situation has ...</p> <p>Got worse, Got better, Or stayed more or less the same?</p>



# The use of midpoints for rating scales

---

I couldn't afford to eat balanced meals?

- 1) Strongly agree
- 2) Agree
- 3) Neither agree nor disagree
- 4) Disagree
- 5) Strongly disagree

# Practical exercise [15 mins]

Review answer scales used in Qs in Questionnaire B

- Are they appropriate? Does the scale match the text of the question stem?
- What kind of data will be produced? What level of measurement?
- Are any answer options missing?
- Are there any changes you would make?

# Designing questionnaires for different modes



# Measurement error and survey mode



# Impact of mode on measurement

	F2F	Telephone	Web	Postal
Information transmission	Both	Aural	Visual	Visual
Interviewer present	Yes	Yes	No	No
Communication channel(s)	Verbal, full paralinguistic, text	Verbal, partial paralinguistic	Text	Text
Interactivity	High	Moderate	Low	None
Cognitive burden	Shared	Shared	Respondent	Respondent

# Reminder of general questionnaire design principles

- Keep questionnaires and survey questions as short and simple as possible
  - Keep Q preambles short
  - Reduce number of words in the Q stem (the bit before the question mark) as far as possible
- Avoid double-barrelled questions, jargon, etc
- Avoid long response lists – cap at 11. If longer lists needed spilt into several Qs or group under sub-headings

# Answers scales & mode

---

## ■ Visual analogue scales

Please put a mark on the line to indicate how you feel today?

Very unhappy \_\_\_\_\_ Very happy

## ■ Verbal answer scales

How happy are you feeling today

(Very happy, happy, unhappy, very unhappy)

## ■ End-labelled answer scales

How happy are you feeling on a scale of 1-10.

1 is very unhappy and 10 is very happy?

# Answer scales

\* End-point problematic in multi-mode as visual presentation varies between modes

	Visual Analogue Scale	Verbal scale (3-4 points)	Verbal scale (5-7 points)	End-point label*	Branched scale
F2F	✓ Separate sheet	✓	✓ Showcard	✓	✓
Telephone	✗	✓	✗	? Rs may forget scale direction	✓
Web	✓ Non response	✓	✓	✓	✓
Paper	✓ Rs may not answer in correct format	✓	✓	✓	✗



# Branched vs. non-branched scales

Branched	Non-branched
<p>How would you rate your overall health?</p> <p>Good</p> <p>Average</p> <p>Poor</p> <p><i>{If Good}</i></p> <p>Would you rate your overall health as Good or Very good?</p> <p><i>{If Poor}</i></p> <p>Would you rate your overall health as Poor or Very poor?</p>	<p>How would you rate your overall health?</p> <p>Very good</p> <p>Good</p> <p>Average</p> <p>Poor</p> <p>Very poor</p>

# Check all that apply (CATA)

- Check all that apply (CATA) **cannot** be used in telephone surveys
- An alternative is the **forced choice** (Y/N) format, where respondents have to answer 'Yes' or 'No' to each item.
  
- Some evidence that the Y/N format better than CATA:
  - Higher levels of endorsement
  - Reduced primacy effects (Smyth et al, 2006; Smyth et al, 2008)
  
- You should not combine CATA and Y/N in a mixed mode study  
However...
- Y/N longer to administer? More burdensome on the respondent?

# CATA vs Y/N

	<b>CATA</b> Code all that apply	<b>Y/N</b>
<b>F2F</b>	✓ Use showcard	✓*
<b>Telephone</b>	✗	✓
<b>Web</b>	✓	✓*
<b>Paper</b>	✓	✓* Rs may omit 'no' response

\* Long lists should be avoided

Source: Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M. and Gray, M. (2013)

# Designing for web: General appearance

---

- Design Qs with **mobile** (small screen) in mind
- Keep images on screen to a minimum: avoid clutter to keep focus on the question
- Headers and logos take up valuable screen space – are they needed?
- Don't include progress bars

# Designing for web: Question stem

- Question text should be visible on a smartphone screen in portrait view. Scrolling is acceptable.
- Important information should be part of the question. Don't assume people will read hidden text or click on information buttons
- Avoid superfluous instructions e.g. tick one only, choose most applicable answer

# Web questionnaire: Answer options

- Grid formats don't work on small screens
- Use collapsible or carousels grids
- Responsive design - format presented to R determined by device (i.e. grid on a laptop, collapsible grid on a smartphone)
- Decide if DK/Ref options immediately visible or are hidden, becoming visible if respondent tries to move on

# Collapsible Grid: Example

Does {#fname} receive additional support at school for any of the following health or behavioural problems or disability? Only consider support that has been offered during this year.

Select one answer on every row.

▾ Mental health problem

yes  
 no

▸ Interrupted schooling

▸ In care of local authority

▸ English as an additional language

▸ More able pupil

▸ Other (please specify)

◀ Previous   Next ▶

Stop ■



Does {#fname} receive additional support at school for any of the following health or behavioural problems or disability? Only consider support that has been offered during this year.

Select one answer on every row.

▸ Mental health problem [yes]

▸ Interrupted schooling [no]

▸ In care of local authority [yes]

▸ English as an additional language [no]

▾ Missing Answer  
More able pupil

yes  
 no  
 Don't know  
 Prefer not to answer

▸ Missing Answer  
Other (please specify)

◀ Previous   Next ▶

# Carousel grid: Example

☰

In general, how much trust do you have in the following types of organisation?

Answer on a scale of 0 to 10, where 0 is 'do not trust at all' and 10 is 'trust completely'.

Banks

0 - do not trust at all
1
2
3
4
5
6
7
8
9
10 - trust completely



Answer on a scale of 0 to 10, where 0 is 'do not trust at all' and 10 is 'trust completely'.

Banks

0 - do not trust at all
1
2
3
4
5
6
7
8
9
10 - trust completely
Don't know

◀ Previous    Next ▶




# Designing for telephone

- Keep total questionnaire length under 30 minutes
- Can questions be read out easily over the phone?
  - Avoid multiple, complex, clauses
  - Avoid / and ( ) e.g. spouse/ partner
- Where possible use short verbally labelled scales (3-4)
- Use Yes/No formats in favour of check all that apply. Use sparingly and avoid long lists of responses.
- Break up long lists into a series of questions

# Question design strategies for mixed mode

---

Start with this approach



## Mode optimisation

Q wording & presentation tailored to individual mode



Exploits strengths of each mode


Enhance overall data quality (De Leeuw, 1992)

### **BUT**

Can introduce measurement differences

## Uni mode

Q wording and presentation is standardised across modes



Minimises measurement differences

### **BUT**

Not all Qs are suitable

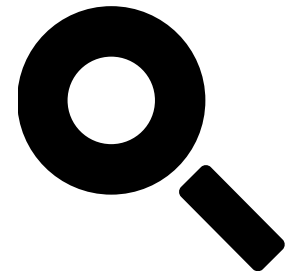
Can increase cognitive complexity

# Practical – 10 minutes

---

## In groups

- Review a question from Questionnaire B for inclusion in a web survey
- Is the Q suitable or are any changes needed?
- What changes would you make?
- If you have time, now review the question for inclusion in a telephone survey. What changes are needed?



Final reflections, questions

# Contact

---

**Debbie Collins**

Senior Research Director

T. 020 7549 7073

E. [debbie.collins@natcen.ac.uk](mailto:debbie.collins@natcen.ac.uk)

Visit us online, [www.natcen.ac.uk](http://www.natcen.ac.uk)

**NatCen** LEARNING  
Social Research that works for society