Cognitive Interviewing and its online implementation “Web Probing”

AAPOR Webinar

Dorothee Behr & Cornelia Neuert, GESIS – Leibniz Institute for the Social Sciences, Mannheim
Presenters

- Head of the Team *Cross-cultural Survey Methods* at GESIS
  - Education: Translation Studies, Diplom equiv. MA; PhD in Translation Studies, University of Mainz
  - With GESIS since 2006
  - Research on cross-cultural web probing and translation of survey questionnaires

- Head of the Team *Questionnaire Design & Evaluation* at GESIS
  - Education: Social Sciences, Diplom equiv. M.Sc.; PhD in Sociology, Uni Mannheim
  - With GESIS since 2012
  - Research on methods for testing and evaluating survey questions and questionnaire design
Web probing is a powerful tool to learn about the response process of respondents and to assess the quality of survey questions.

Web probing studies need to be carefully designed to ensure response quality.

Web probing does not replace cognitive interviewing; rather, it should be seen as a useful supplement to cognitive interviewing and other pretesting methods.

As other methods, it has its strengths and weaknesses.
Overview

- Basics: Key aspects of Cognitive Interviewing
- Web Probing
  - Strengths and weaknesses
  - Probing techniques
  - Design of probes
  - Respondent recruitment
  - Implementation stages
  - Use scenarios
  - Analysis
Cognitive Interviewing

“…entails administering 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 its author intends.”

(Beatty & Willis, 2007, p. 288)
The aim of cognitive interviewing is to get insights into the cognitive processes underlying survey responding:

- How do respondents interpret questions?
- How do they retrieve relevant information and events from memory?
- How do they arrive at a judgment about what to answer?
- How do they map their “internal” answer to the answer categories provided?

(Courangeau et al., 2000)
Conducting Cognitive Interviews

- No clear/fixed rules
- Practice ranges from (completely) unstructured to (completely) standardized interviewing
- No clear evidence of how many interviews should be conducted per pretest
- Usually small number of interviews (cases)
- Test persons should resemble the respondents of the actual survey with regard to sex, age, education, and other important study-specific variables
- Duration of max. 90 min
- Instruction: Not you are being tested, but the questionnaire
- Test persons receive a monetary compensation for participating
Example

Question:
“I feel more like a citizen of the world than of any other country.”
Strongly agree/ Strongly disagree (5-point scale)

Findings:
- Only 2 out of 20 test persons say that they don’t know (or that they are unsure of) what the phrase “citizen of the world” refers to
- 18 select an answer category (without hesitation)

(Lenzner et al., 2013)
Findings:

▪ Cognitive pretesting techniques: 5 out of 20 test persons interpret the phrase wrongly in the sense of “human being,” for example:
  – “For me, citizen of the world refers to all human beings living on earth.”
  – “Everyone who lives on earth is a citizen of the world.”

▪ All 5 respondents “strongly/rather agree” with this statement

⇒ Proportion of people feeling like “citizens of the world” in a population would probably be overestimated considerably with this item.

(Lenzner et al., 2013)
Probing

- Probing is a technique that uses follow-up (or “probing”) questions administered either immediately after the respondent provided an answer or at the end of the interview.

- The goal is to gather specific information about respondents’ understanding of terms, questions or answer categories and about the processes by which they arrived at their answers.
Probing

Variants of probing:

- Comprehension probing
- Category selection probing
- Information retrieval probing
- Confidence rating
- General/elaborative probing
Web Probing

Target Question

Have you used any of the following devices or systems that are connected to the internet? Tick all that apply or k).

a) Solutions for energy management for your home, such as internet-connected thermostat, utility meter, internet connected lights or plug ins
b) Solutions to ensure the security/safety of your home, such as internet-connected home alarm system, smoke detector, security cameras, door locks
c) Home appliances such as internet-connected robot vacuums, fridges, ovens, coffee machines
d) A virtual assistant in the form of a smart speaker or of an app, such as Google Home, Amazon Alexa/Echo
/Computer, Google Assistant, Siri, Cortana, Bixby
e) Wearable devices connected to the internet, such as smart watches, fitness bands, connected goggles or headsets, connected accessories, clothes or shoes
f) Internet-connected devices used for health and medical care, for example, devices for monitoring blood pressure, sugar level, body weight (e.g., smart scales)
g) Entertainment devices connected to the internet, such as smart TVs, smart speakers, game consoles etc.
h) Toys connected to the internet, such as, robot toys (including educational) and dolls
i) A car with built-in internet connection
j) Other devices not mentioned above:
k) I have not used any connected devices or systems

Open-ended probe

We would like to get some more information about one of the previous questions. The question was: “Have you used any of the following devices or systems that are connected to the internet?”

Which wearable devices did you have in mind when answering this question? Please describe these in more detail.

Closed probe

We would like to get some more information about one of the previous questions. The question was: “Have you used any of the following devices or systems that are connected to the internet?”

You indicated that you have used a car with built-in internet connection.

Was this car equipped with internet access (e.g. in-car wifi) or did you connect to the internet via a smartphone/tablet?

- Car was equipped with internet access
- I connected to the internet via a smartphone/tablet

(Lenzner et al., 2018)
Web Probing - Example

For each of the following statements, please select the response which best describes your work situation.

- Your colleagues help and support you
- You are consulted before objectives are set for your work
- You are involved in improving the work organisation or work processes of your department or organisation
- You have a say in the choice of your work colleagues
- You can take a break when you wish
- You have enough time to get the job done

(Hadler et al., 2018)
Probing Variants

- General probing
  
  The previous question asked whether your colleagues help and support you. Your answer was "Most of the time". What were you thinking of when answering this question? Please give us some examples.

- Specific probing
  
  The previous question asked whether you are consulted before objectives are set for your work. What sort of work objectives were you thinking of in this question?
Probing Variants

- **Category selection Probing**

  The previous question asked whether you have a say in the choice of your work colleagues. Your answer was "Never". Why did you select this answer?

- **Comprehension probing**

  Who are you thinking of, when you read "colleagues" in this question?
Probe wording

- Requires specific attention ("one-shot")
- Be precise and specific
- Avoid addressing more than one topic
- The depth and length should be clear (key words, small essay, etc.) depending on probe type
Strengths

- **Sampling & recruitment**
  - Larger samples
  - Representative samples (region)
  - Recruitment time- and cost-efficient

- **Mode**
  - Higher standardization across subjects, self-administration rules out interviewer effect
  - No transcription needed

- **Analysis**
  - Findings can be quantified
  - Rules out ‘false positives’

(Behr et al., 2017; Lenzner & Neuert, 2017)
Weaknesses

- **Mode**
  - Lack of motivating interviewer
  - No spontaneous or conditional probing
  - Only scripted probes possible, no follow-up on incomplete answers (“one-shot”)

- **Response quality**
  - Higher amount of “probe nonresponse” and mismatching responses
  - Shorter answers, not interpretable answers
  - Higher response burden

- **Analysis**
  - Higher effort in data analysis due to larger sample size

(Behr et al., 2017; Lenzner & Neuert, 2017)
Comparability of Methods

But most importantly:

Web probing and f2f cognitive interviewing detect very similar problems and lead to the same suggestions for item revisions

Comparability - Example

Question: *How important is it that citizens may engage in acts of civil disobedience when they strictly oppose government actions?*

- Probes: What does the term “civil disobedience” mean to you?/Can you explain your answer a little further?

- Problems identified:

<table>
<thead>
<tr>
<th></th>
<th>f2f</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term ‘civil disobedience’ is unfamiliar/undefined</td>
<td>30%</td>
<td>5%</td>
</tr>
<tr>
<td>‘Civil disobedience’ is associated with violent behavior</td>
<td>15%</td>
<td>12%</td>
</tr>
<tr>
<td>The response scale is interpreted as reaching from nonviolent to violent behavior</td>
<td>5%</td>
<td>2%</td>
</tr>
</tbody>
</table>
Design of web probes
Probe placement

- Place a probe on a separate screen to keep the response process for the closed question distinct from the response process for the open-ended probe.
- Similar to face-to-face cognitive interviews
Probe presentation

- Provide needed information on the probe screen to reduce respondent burden, e.g. the corresponding closed question and/or the response option that a question refers to.

Note: Use of placeholders in the survey.

The question was: "And how important is it that people convicted of serious crimes lose their citizen rights?"

Your answer was "3" on a scale from 1 (not at all important) to 7 (very important).
Sequence of probes – within a survey

- Repeated identical probes (e.g., category selection probe with same layout and text box size) lead to habituation effects among respondents; respondents may not read probe text of different probes, if the overall layout and text box size is the same.
- Therefore, carefully reflect on the sequence and design of probes.
Sequence of probes – multiple probes for one question

- First evidence suggests that a category selection probe should come first, before other probes, such as comprehension and specific probe:
  - Increase of response rate
  - Decrease of “mismatching” answers (those that do not fit the probe asked)

(Meitinger et al., 2018)
Text box size

- Adapt the size to the expected answer
Legal immigrants are important, if an individual can not obtain legal status, or refuses to, it puts strain on the economy of migrant workers, service industries. All groups, none excluded.
21 probes (for a 20-25 min) in a survey are possible.
Slightly higher break-off rate with more probes, may require over-sampling, but no differences with nonresponse, number of themes, etc. in experimental study.

(Neuert & Lenzner, 2021)
The study referred to used online access panelists; these were paid; the respondents were informed on the start screen of the survey that this was a pretest and that answers to the open-ended questions would help to revise survey questions.

Alternative experiences: Other web probing studies without explicit announcement of pretesting study worked well with 6 – 8 probes in a 15-minute survey. (Behr et al., 2017)
Nonresponse

- Open-ended questions in general are prone to nonresponse and so are cognitive probes.
- Different types of nonresponse are shown below, but nonresponse also depends on the research question of the researcher (e.g., Are 1-word answers OK? Is a don’t know answer helpful?)

<table>
<thead>
<tr>
<th>Category</th>
<th>Type of Probe Nonresponse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>Complete nonresponse: respondent leaves a text box blank</td>
</tr>
<tr>
<td>Category 2</td>
<td>No useful answer: response is not a word e.g., “dfqjih”</td>
</tr>
<tr>
<td>Category 3</td>
<td>Don’t know: e.g., “I have no idea,” “DK,” “I can't make up my mind”</td>
</tr>
<tr>
<td>Category 4</td>
<td>Refusal: e.g., “no comment,” “see answer above”</td>
</tr>
<tr>
<td>Category 5</td>
<td>Other nonresponse: responses that are insufficient for substantive coding: e.g., “my personal experience,” “it depends,” “just do,” “just what it is”²</td>
</tr>
<tr>
<td>Category 6</td>
<td>One word only: respondent just writes a single word, e.g., “economy”³</td>
</tr>
<tr>
<td>Category 7</td>
<td>Too fast response: respondent takes less than two seconds to answer</td>
</tr>
</tbody>
</table>

(Behr et al., 2017)
How to tackle nonresponse I

- Appropriate wording and design of probes (see slides before).
- Use of soft checks in survey software that “gently” remind respondents of an answer but that eventually allow respondents to move on without an answer.
Automatic checking of nonresponse and tailor-made follow-up probe with motivational sentence:

- Video zu EvalAnswer: http://kaczmirek.de/webprobing/video
- Open-access Tool (source code): https://git.gesis.org/surveymethods/evalanswer
- Languages: English, Spanish, German
- Can also be used for nonresponse coding after data collection
Closed probes vs. open-ended probes

- **Closed probes:**
  + Quantification of patterns of interpretation and errors
  + Reduced costs and burden during data analysis
  + Easier for respondents
  + Less probe nonresponse

- Both probe formats do not provide comparable results:
  coverage of themes, patterns of interpretation, number of themes

(Neuert et al., 2021; Scanlon, 2019; 2020)
Closed probes vs. open-ended probes

- Select an appropriate approach for own research question:
  - Closed probes are useful when the researchers have a particular *hypothesis* they would like to investigate,
  - … and the *number* and *type* of response categories can be determined in advance.
  - Open-ended probes are useful when the objective is more general or if one is interested in *full breadth* of interpretations.

(Neuert et al., 2021; Scanlon, 2019; 2020)
Respondent recruitment
Access to respondents: access panels

- Non-representative panel – respondents who have voluntarily signed up for taking part in surveys in regular intervals; information on panels …
  - … in 28 Questions to Help Buyers of Online Panels and/or in so-called panel books.
  - Panels may also be ISO certified - ISO 20252:2019: Market, opinion and social research, including insights and data analytics -- Vocabulary and service requirements
  - Useful reference: AAPOR report on Online Panels
  - Example studies: Hadler, 2021; Meitinger & Behr, 2016; Neuert & Lenzner, 2021
Crowdsourcing – the practice of turning to a large group of people to obtain work, information, or opinions.

Crowdsourcing platforms, e.g.:
- Amazon Mechanical Turk (MTurk)
- [Facebook may be used in a similar vein]
- Example studies: Edgar et al., 2016; Fowler & Willis, 2020
Access to respondents: probability-based panels

- Existence of such panels based on availability in a country
- Open Probability-based Panel Alliance:
  - [https://openpanelalliance.org/](https://openpanelalliance.org/)
  - Panels in the US (*Understanding America Study*), Germany (*GESIS Panel*), South Korea (*Korean Academic Multimode Open Survey*), and the Netherlands (*LISS Panel*)
- However, it may depend on panel policy to what extent open-ended questions are possible.
Access to respondents: own resources

- Proprietary panels or own respondent pool
Implementation stages
Implementation ...

- **... at the pretesting stage** to assess questions prior to a study:
  - Alone or in combination with traditional cognitive interviews (see next slides for an example).
- **... at the main production stage** in a web survey to learn about the validity of the actual survey questions:
  - Schuman’s random probe method may be an option (1966), whereby splits of respondents receive probes for selected questions.
- **... for post-hoc evaluation** (see next slides):
  - To explain anomalies in the data or to assess re-occurring questions that are deemed problematic in general.
Use scenarios
Combining methods - Example

- Project: Cognitive pretesting of the EWCS questionnaire from wave 6.
- Testing aim:
  Are there differences in how employees, as compared to self-employed respondents, understood items that pertain to job quality?

(Hadler et al., 2018)
Combining methods - Example

Task: Job quality survey results

Web Probing (3 countries)

Comprehension and category selection of job quality indices

F2f-Interviews (2 countries)

Deeper examination of results from WP

(Hadler et al., 2018)
Combining methods - Example

**Task**: Job quality survey results

- **Web Probing** (3 countries)
  - Comprehension and category selection of job quality indices
- **F2f-Interviews** (2 countries)
  - Deeper examination of results from WP

**Target Question**: Your colleagues help and support you. 5 point scale: Almost – Most of the time – Sometimes – Rarely - Never

**Web Probing**: Additional response option “not applicable”

“not applicable”-responses:
- Self-employed: 33% – 55%
- Employed: 3% – 5%

**Face-to-face Interviews**: Mimic of interview situation without “not applicable”

- Self-employed included other self-employed people working in the same field or referred to their employees

**Probe finding**: Respondents who chose “not applicable” had no colleagues to whom they could refer this statement.

(Hadler et al., 2018)
Cross-cultural web probing

- Sound method to study equivalence of items and identify different answer patterns across countries.
- Important questions to address:
  - Coding in original language or based on translation of open-ended answers (translation guidelines and challenges in Behr, 2015; Dorer et al., 2021)
  - Code scheme development taking into account sample answers from all countries/cultural group to prevent biased coding schemes.
Example “Social security system”

- Post-hoc evaluation, combining measurement invariance test and web probing.
- Topic: National Identity (items from ISSP 2013 module)
  - How proud are you of [COUNTRY] with regard to its social security system?
  - ¿Qué tan orgulloso/a está Ud. de México con respecto a su sistema de seguridad social? (→ plus: implementation of specific probe in web probing study)
  - 39% of web probing respondents were thinking of the security situation in Mexico.

(Meitinger, 2018)
Combination of Methods

- Cognitive Interviewing
- Web Probing
- Web Probing
- Cognitive Interviewing
- Web Probing
- Web Probing
- Cognitive Interviewing
- Web Probing
Analysis
Errors or themes?

- Coding of **errors** based on (established) error coding schemes, for instance, ‘information unavailable’ or ‘problematic term’.
- Coding of **themes** (inductive/deductive approach), for instance, ‘literature’, ‘music’, ‘performing arts’ when probed: “What particular achievements in the arts and literature did you have in mind when you were answering the questions?”
Wrap-up
Web probing is a powerful tool both for substantive research and to evaluate questions (comprehensibility, validity & comparability).

The implementation (wording and layout) determines the answer quality.

Further methodological research is needed to ensure high response quality.
Thank you for your attention!
References (slides and beyond)


Web probing - methods research II

Web probing - substantive research


- Braun, M., & Johnson, T. P. (2018). How should immigrants adapt to their country of residence?: A mixed methods approach to evaluate the international applicability of a question from the German General Social Survey (ALLBUS). In E. Davidov, P. Schmidt, Jaak Billiet & B. Meuleman (Eds.), *Cross-cultural analysis: methods and applications* (pp. 615-632). Routledge.


Cognitive interviewing and response process:


GESIS pretests involving web probing:


Voice recording

Further examples: left vs. right

<table>
<thead>
<tr>
<th>Poor views, uniformed</th>
<th>making informed, unified decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very liberal, sometimes overcompensating for things that need changing.</td>
<td>Very conservative, not changing even in light of blatant need.</td>
</tr>
<tr>
<td>Lies, slander, despotism, hypocrisy and manipulation of the weak willed.</td>
<td>Charitable, honest, hopeful, and hardworking.</td>
</tr>
<tr>
<td>Concerned with people's rights</td>
<td>Concerned with corporate rights</td>
</tr>
<tr>
<td>Liberal, government social programs, pro-choice, more government control</td>
<td>Conservative, hopefully less government spending, pro-life. Capitalism</td>
</tr>
<tr>
<td>Democrat, liberal views</td>
<td>Republican, conservative views</td>
</tr>
<tr>
<td>Socialist view, more regulations</td>
<td>Less interferences of governments, individual is more self reliant</td>
</tr>
<tr>
<td>Liberal, open minded to new ideas.</td>
<td>Conservative, RICH, closed minded to any change that may help the average or poverty level of the general population.</td>
</tr>
<tr>
<td>Helping the poor</td>
<td>Big business making profits</td>
</tr>
<tr>
<td>Do not have any universal healthcare, big government, own control</td>
<td>Do not have any small government, fair tax, right to bear arms</td>
</tr>
<tr>
<td>My environmental opinions, right to choose. right to die. etc.</td>
<td>Immigration policies, taxes.</td>
</tr>
<tr>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>-66</td>
<td>-66</td>
</tr>
</tbody>
</table>
Further examples: individual solutions

Source: CICOM

| pidlr1: "Please specify what type of individual solutions you had in mind. What factors should determine the ideal role distribution in a family?"
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>It depends on the needs of the family. If the breadwinner (husband or wife) can support the family, then one works and the other stays home, or if need be, both work.</td>
</tr>
<tr>
<td>The location the family lives, the lifestyle they want, the income bringing in.</td>
</tr>
<tr>
<td>The ideal role in a family should be determined by practical application. For example, if the woman has higher education and more work experience then it seems logical that she would be the primary source of income in a family.</td>
</tr>
<tr>
<td>The man or women who has the greatest earning potential should work to financially support the family. The non-primary earner should work either inside the home or outside the home to support the family.</td>
</tr>
<tr>
<td>the two people should sit down and discuss their opinions, feeling etc. and both agree how role distribution should be; but be open to reassessing their own particular situations.... things change; as do people.</td>
</tr>
<tr>
<td>Every family is different and they should do what works for them.</td>
</tr>
</tbody>
</table>
Further examples: satisfaction with democracy

Source: CICOM
Effects of probing?

- Couper (2013) provides evidence that additional “commenting” (explaining or clarifying responses) can have an effect on answers to sensitive closed questions
  - Comment possibility decreased the level of prejudice reported in a set of 10 items on attitudes towards immigrants
- Fowler and Willis (2020) compare concurrent and retrospective probing
  - Some limited evidence for differences in responses to closed items (though not large)
- Scanlon (2019) compares the effect on survey response of embedded closed-ended probes (vs. no probes at all)
  - No overall negative impact