Questionnaire design issues: asking the right questions

Jettie Hoonhout
Philips Research
November 2013
- Background in Human Factors Psychology
- University of Utrecht, University of Maastricht
- Philips Research Eindhoven since 2000
Overview

- Issues in the design of your own questionnaire / survey
- Issues in the use of existing scales
- Slash upgrade existing questionnaire
- Work through a number of questionnaire items
subjective assessment:

indirect “observation” techniques →
collecting information from users about their experiences and opinions →
the users of a system that is being studied act as measuring instruments

(but instruments might have imprecisions and give inaccurate readings....)
Subjective assessment formats

• Face-to-face interview (unstructured, semi-structured, structured)
• Telephone interview
• Self-administered questionnaire (mail, website, ..)

• Questionnaires versus scales:
  – Questionnaires to address specific research questions in the context of a particular project (Use and appreciation of device X in the context of project Y)
  – Scales to conduct repeated measurements across different studies to determine some psychological construct: e.g. satisfaction, pleasure, engagement, presence
when (not) to use....

several performance indicators are better determined through observation (e.g. task completion, time on task, errors, use of support tools (manuals), effort...)

other performance indicators can not so easily be determined through observation (perceived ease of use, perceived comfort, attitudes, opinions, motivations,...)
Questionnaire design requirements:

- Precision/sensitivity/discriminate
- Reliability: results are stable across multiple administrations of the instrument
- Validity: The questionnaire measures what it says it is measuring
- Usability to respondent and researcher (designing a questionnaire is like designing a product!)
Writing questions should not be the first thing on your mind!

• Carefully consider research questions and objectives

• Carefully consider how to conduct the analysis

• And even before that: carefully consider if a questionnaire is the right instrument given project questions and objectives!
questionnaire design considerations:

1. questionnaire purpose ← → research aim
2. respondent selection (non-response)
3. question content
4. question wording
5. response format
6. question sequence
7. questionnaire length
8. questionnaire layout
9. **piloting** (= testing a questionnaire, e.g. talk aloud, cognitive interviewing)
completing a questionnaire has both social and cognitive components:

• social encounter, comparable to conversation, following rules for conversation
• cognitive task for respondent, tapping into language understanding, memory, reasoning, etc.
### question-answer process

<table>
<thead>
<tr>
<th>question</th>
<th>1. interpretation of question and response options</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. retrieval of information from memory</td>
</tr>
<tr>
<td></td>
<td>3. formulating the response, and checking the response against internal standard</td>
</tr>
<tr>
<td></td>
<td>4. matching response with answer categories, editing and reporting the answer</td>
</tr>
</tbody>
</table>
some potential pitfalls…

- selection of respondents
- failing memory of respondents
- effect of fatigue, feeling bored, time pressure
- social desirable replies
- respondent does not want to appear stupid
- “please the researcher” effect
closed-ended questions:

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy and quick to answer</td>
<td>Can put ideas in respondent’s head</td>
</tr>
<tr>
<td>Answers across respondents easy to compare</td>
<td>Respondents with no opinion might answer anyway</td>
</tr>
<tr>
<td>Analysis of answers easier</td>
<td>Respondents can feel constrained and frustrated</td>
</tr>
<tr>
<td>Response choices make question clearer</td>
<td>Many choices can be confusing</td>
</tr>
<tr>
<td>Easy to replicate study</td>
<td>Cannot tell if respondent misinterpreted the question</td>
</tr>
<tr>
<td></td>
<td>Fine distinctions may be lost</td>
</tr>
</tbody>
</table>

Philips Research, November 2013
# open-ended questions:

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>permit unlimited number of answers</td>
<td>respondents give answers with different level of detail</td>
</tr>
<tr>
<td>respondents can qualify and clarify responses</td>
<td>answers can be irrelevant</td>
</tr>
<tr>
<td>can find the unanticipated</td>
<td>inarticulate or forgetful respondents are at disadvantage</td>
</tr>
<tr>
<td></td>
<td>coding responses is subjective and tedious</td>
</tr>
<tr>
<td></td>
<td>requires more time and effort from the respondent</td>
</tr>
</tbody>
</table>
Some tips

• a questionnaire with a title is generally perceived to be more credible than one without
• begin with a couple of non-threatening and interesting items
• leave ample white space; for comments, but it also makes the questionnaire appear easier
• test online questionnaires in different browsers, different screen sizes
• the affiliation of the researcher can make a world of difference
• resist the temptation to add more and more questions….
Finally…. Writing questions…
Question wording

• Use simple wording (but not too simple…)
• Be brief (although questions should not be too short – reduces credibility)
• Be specific
• Avoid:
  – Being condescending or talking down to respondent
  – Using biased wording
  – Using abbreviations or scientific jargon
• Questions should ask for only 1 piece of information, so avoid:

• Asking two questions at once
  – "Do you buy frozen meat and frozen fish?" A "Yes" answer can mean the respondent buys meat or fish or both.

• Asking questions that contain assumptions
  – Did you like product X? (assumption that respondent has bought at least once this product)
• A questionnaire gets people to express their feelings, perceptions, behaviors, and experiences, both past & present; but do not overask, and avoid:

• Asking hypothetical questions
  – If we are going to build in feature x, are you going to use it?
• Asking for solutions to complex issues, or issues about which a respondent most likely never thought of before
• Question wording should ensure that every respondent will be answering the same thing, so avoid:

• Ambiguous wording or wording that might mean different things to different respondents
• Using terms for which the definition can vary (If it is unavoidable, provide the respondent with a definition.)
• Being ambiguous about the time period the respondent should consider
On scale points
not at all 0---0---0---0 very much

• A good question is also one that elicits a range of responses.
• Two or three (realistic) options may be appropriate, but four, five or six will usually produce a more interesting result, especially as respondents may avoid the extremes.
• Better label the extreme responses in a 'mild' way for this reason. Use 'poor' rather than 'bad' and 'very good' rather than 'excellent'.
• Even / non-even number of points on a scale?
• 5, 7, more, ..., ?
(Re)usage of existing scales

Self assessment manikin
Scales used in hair removal studies

• Focus on pain, with large parts of the scale referring to severe levels of pain → not really the business we are in...

• Not full use of scale points
  – Psychometric issues

• Pain is a psychological construct
  – Not all pain is the same, depends on various factors, such as, ‘purpose of application inflicting the pain’, test conditions

A number of examples of pain scales, from well-known validated ones (e.g. VAS, top right corner), to more funny ones (bottom ones)
Scale design requirements:

• Precision/sensitivity/discriminate
• Reliability: results are stable across multiple administrations of the instrument
• Validity: The scale measures what it says it is measuring
• Usability to respondent and researcher (a scale is like a product!)
Before adopting a scale, check a range of properties (Green, Dunn & Hoonhout, 2008)

1) Construct definition
- **Aim of construct definition**: This is arguably the most important element of scale development: good scale items cannot be formulated without it. Determine what you are and are not intending to measure.
  - 1a) Is the construct grounded in theory?
  - 1b) Clarity: need to know what you measure is not ambiguous and confused, when it is you are unsure of what you are measuring.
  - 1c) Discriminating: you must know that you are not measuring something that you are not intending to measure (confound).

2) Scale validity
- **Aim of scale validity**: The extent to which a scale measures what it intend to measure.
  - 2a) Construct validity
  - 2b) Context
  - 2c) Sample

3) Scale reliability
- **Aim of scale reliability**: Ensure that the measure consistently reflects the construct, not only internally (Cronbach’s alpha) but also over time (test-retest).
  - 3a) Inter-item reliability
  - 3b) Test-retest reliability
Validity of scales

• Refers to measuring the construct of interest, the whole construct, and nothing but the construct

• But constructs often ill defined!

• Dutch Psychologist Piet Vroon: “IQ is what is measured by the test”
Construct validity: “refers to measuring the construct of interest, the whole construct, and nothing but the construct”

- (Face validity – does it appear to do what it promises to do?)
- Content validity – representative set of items for domain of interest; does it cover all aspects of the construct
- Criterion validity – relation between scores and other, independent criterion (variables, e.g. particular behavior)
  - Predictive validity
  - Concurrent validity
- Convergent validity – correlation between this measure and e.g. other, alternative scales
Discussion

- Scales developed for certain settings might not be applicable in others (e.g. mood scales such as PANAS, POMS)
- Underlying construct quite often not fully clear, underlying theory still under development
- Published scales often claim to be valid, but might turn out not to be adequately tested…
- If internal consistency is high (cronbach’s alpha) certainly something is measured, but not yet clear what!!
- Apart from validity: norms…. 
Issues I often encounter

- Using a questionnaire when one should not
  - E.g. because one is afraid to use interviews instead, or other means
- Not resisting to ask many, too many questions
- Thinking that respondents are as much interested in your project as you are
- Using lots and lots of words that are familiar to the researcher
- Using scales in a usability test with say just <20 participants
- Using scales in a usability test without anything to compare the results with
- Not pilot testing a questionnaire
- No proper pre-investigation before developing the questionnaire
references:

- no bedtime reading, but essential to understand the limitations of asking questions: Nisbett, R.E. and Wilson, T., Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 1977, 84(3), 231-259.
Group assignments

1. Work through the examples of questionnaire wording: what do you think might be the impact on the answer a respondent is going to provide, how would you perhaps change the wording, and why

2. Review the existing questionnaire, and come with suggestions for improvement.
QUESTIONNAIRE DESIGN:
TWO EXAMPLES
Two examples of questionnaire/scale: Home video editing & Involvement with media

- Questionnaires versus scales:
  - Questionnaires to address specific research questions in the context of a particular project (Use and appreciation of home video editing in the context of project EditWhileWatching)
  - Development of Involvement Scale to conduct repeated measurements across different studies to determine the level of involvement in studies of audio/video quality and quality disturbances
Appropriate choice....

Difficult to observe home video editing practices and use of current tools for that;

Involvement might be determined e.g. through secondary task measures, but more complex to “administer”; also, assumed that it is multidimensional construct, which will be more difficult to inspect with secondary task measure

In both cases: opinion and the perception of the respondent of the situation is important
Global outline of development process in both cases (not all steps applicable for both cases)