Survey design - best practice guidelines

This document outlines some of the main considerations in survey question and answer design.

Key considerations for designing individual questions include:

1 brevity

2 clarity and consistency

3 non-leading

4 prestige bias

5 double-barrelling

6 consistency of phrasing

7 answer categories

8 frequency questions

Brevity

You should design your questions to be short and easy to understand. And use plain English where possible. This will allow the respondent to complete the questionnaire quickly. They should understand what is being asked on the first reading.

A quick test is to ask a lay person to read the question; if they do not understand it the first time you should simplify the question.

For example

Q How would you rate the quality of the educational supervision you received during the whole time you were on this course?

This question is long and uses words that are not necessary for the question to be understood or answered.

In this case the question should be shortened as follows.

Q How would you rate the educational supervision during this course?

Clarity

Your questions should be clear, succinct and unambiguous. The goal is to eliminate the chance that the question will mean different things to different people. To this end, it is best to phrase questions empirically if possible and to avoid the use of unnecessary adjectives.

A key standard for a good question is that all the people answering it should understand it in a consistent way and that it should mean to them what the researcher expected it to.

Ambiguous words such as 'usually' or 'frequently' have no specific meaning and need qualifying. You should avoid using them in your questions.

For example

Q Do you usually read a newspaper?

A Yes/No

One person could answer yes to this question and read a newspaper once per week, while someone else who reads a newspaper everyday would also answer yes to this question.

This question should be reworded as follows.

Q On how many days during the last seven, if any, have you read a newspaper?

A Everyday, four to six days, one to three days, none

This quantifies the time period to be covered and also the answer categories.

A common reason for questions not being clear is asking about too many aspects in the one question. This is covered in double barrelling below.

Non-leading

A leading question is one that forces or implies a certain type of answer. The way a question is phrased can make it leading as most adjectives, verbs, and nouns in English have either a positive or negative connotation.

You should check each question for leading phrases.

For example

The use of wording such as 'good', 'be able' and 'enabled' can be leading.

Q How good was this course?

Regardless of the answer categories the question wording implies that the course was good.

The question should read:

Q How good or poor was this course?

This balances the question and does not imply a positive or negative response

A common way in which leading questions are used in surveys is with the phrase 'how many' or 'how often'.

Q How many times have you worked over eight hours a day, in the last five days you have worked?

OR

Q How often have you worked over eight hours a day, in the last five you have worked?

The 'how many' or 'how often' implies that the respondent has worked late in the last five days. If they have not worked late the question implies that the researcher expects that they have, so they are more likely to conform and answer that they have worked late even if this is not correct (this could also introduce prestige bias, see below).

The question wording could be altered to the following.

Q How often, if at all, have you worked over eight hours a day, in the last five days you have worked?

Prestige bias

Prestige bias is the tendency for respondents to answer in a certain way to conform to social norms or peer groups.

The above question is a good example of prestige bias in that trainees may assume that part of the trainee role is working long hours.

Double-barrelling

If the wording of a question can be broken down into two or more separate questions, and a respondent could answer those separate questions in different ways, then it will be impossible to analyse the responses to the combined question without ambiguity.

You should aim to ask about only one variable per question.

For example

Q Do you agree or disagree that patient handover at shift change took place safely and reliably?

A Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree, don't know.

This is a double-barrelled question. The handover may have been reliable, in that it happened at the same time each day and the same people were involved, however the trainee may think it was not particularly safe as handover was discussed with junior staff for example.

The question should be split into two questions, asking each variable in a separate question.

Consistency of phrasing

It's better to stick to consistently phrased statements throughout for people to respond to either positively or negatively. If you switch between positive and negatively phrased statements respondents could get confused and may tick a response option they didn't mean to.

For example

Q The training course increased my understanding of medical education (the positive response would be strongly agree / agree)

Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree

Q The training course left me concerned about gaps in my knowledge (the positive response here would change to disagree / strongly disagree)

Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree

Q The training course was delivered at the right pace for me (the positive response would be back to strongly agree / agree)

Strongly agree
Agree
Neither agree nor disagree
Disagree
Strongly disagree

Answer categories

The design of your answer categories is just as important as the design of your question. Answer categories can introduce bias, be unclear and leading, and generate invalid data.

For example

A closed format question must supply answers that not only cover the whole range of responses, but that are also equally distributed throughout the range. All answers should be equally likely. An example where the answer categories are not equal would be a question that supplied these answer choices :

- superb
- excellent
- great
- good
- fair
- not so great

Balancing a scale with the same number of positive and negative responses is very important; limiting the scale as shown above to mainly positive responses will introduce bias.

This scale should be:

- very good
- good
- neither good nor poor
- poor
- very poor
- no opinion

In self-completion attitude questions, Dillman (2000) recommends an even number of response categories, with the last one being a 'don't know/no opinion' option and a 'neutral' option as the central value of the remaining (uneven number) options. Labelling all of the values and not leaving choices to be inferred from their place in the overall pattern is also recommended.

With all semantic scales, the wording of the 'anchor statements' (the statements at the end of the scale) is crucial to the distribution of data that is likely to be achieved. A five point bipolar scale that goes from 'extremely satisfied' to 'extremely dissatisfied' is likely to discourage respondents from using the end points and to concentrate the distribution on the middle three points. If the endpoints were 'very satisfied and 'very dissatisfied' the end points would be used by more respondents and the data would be more widely distributed across the scale.

This can make the data more discriminatory between items. As a general rule, the stronger the anchors, the more points are required on the scale to obtain discrimination.

Frequency questions

When asking about the frequency of a variable, you should quantify the scale and keep it on the same continuum.

For example

When asking a question about frequency, rather than supplying choices that are open to interpretation such as:

- very often
- often
- sometimes
- rarely
- never

it is better to quantify the choices, such as:

- every day or more often
- two to six times a week
- once a week
- less than once per week
- never

Scales should never be mixed. The scale below uses a mixture of quantifying scales (daily, weekly, monthly, never) and those open to interpretation (rarely). It is also not on a continuum as you would expect with options in a timescale:

- daily
- weekly
- monthly
- rarely
- never

If someone wants to answer every two weeks for example this scale does not offer this option. Timescales should be mutually exclusive and on a continuum.

You should test every question in a questionnaire or survey to ensure that none of the factors identified above are contained within the question or answer categories. This will improve the reliability and validity of your data.

The factors highlighted above demonstrate the most common errors in question and answer design, however, this list is not exhaustive.

Remember

Brevity
Keep your questions short. Cut out any unnecessary words.

Clarity
Avoid ambiguous language including words such as 'usually' or 'frequently'.

Non-leading
Avoid using leading phrases such as 'how often' or 'how good'.

Prestige bias
Be aware that some respondents will exaggerate where they think a response is desirable (such as working long hours).

Double-barrelling
Don't ask more than one variable in a question.

Consistency of phrasing
Stick to positively phrased statements throughout

Answer categories
Make sure your response categories are balanced and include all the possible responses, avoid strong anchor statements 'extremely satisfied'.

Frequency categories
Make sure your response categories are quantified, mutually exclusive and on a continuum.

Contact information

Rebecca Barnard (Survey Design Manager, Surveys Team)

rebecca.barnard@gmc-uk.org