

Response alternatives

So far, we have been discussing requests for answers. As was indicated in Chapter 3, the requests can have many different forms, which in turn can create the same response alternatives for the respondent. However, the fact that the same response possibilities are present does not mean that the requests for an answer measure the same thing. Along the same line, it is not immediately clear whether requests for an answer that are identical but differ in the set of possible responses measure different variables. This is an empirical question which has to be answered for different measures. Saris (1983) showed that at least some sets of response scales, although different, will give responses that are identical, except for a linear transformation suggesting that roughly speaking, these measures are indeed identical.

Another issue studied by many people is whether it makes sense to present the respondents with more than only a few categories. Most textbooks suggest, in reference to Miller (1956), that people can not use more than approximately 7 categories. Cox (1980) has argued that Miller's rule does not apply at all to this problem. He suggests that more information can be obtained if more categories are used. This opinion is shared by a few more researchers (Saris et al. 1977; Andrews 1984; Alwin 1997; Köttringer 1995).

Finally there are people who suggest that it would be advisable for certain problems [Krosnick and Fabrigar (forthcoming)] or in general in qualitative research, not to use explicit response alternatives. They suggest that requests with open answer categories are the best because they do not force the respondents in the frame of reference of the researcher.

All these options will be discussed below. The arguments pro and con will be mentioned and an empirical evaluation of the effects on data quality of the different possibilities will be given in Part III of this book.

5.1 OPEN REQUESTS FOR AN ANSWER

As has been mentioned above, some people argue that requests with open answer categories are better than requests with closed categories because people can follow their own thoughts and are not forced in the frame of refer-

ence of the researcher. A request that is exemplar for this dilemma and which has been studied frequently is as follows:

5.1 *What is the most important problem that our country is confronted with nowadays?*

This request can be asked as an open request as indicated above or with possible responses, chosen on the basis of prior research based on the open request. A comparison between these two requests has been studied several times by Schuman and his colleagues. Schuman and Presser (1981) reported that the results from the two requests are very different. The open request seems to be influenced by events that were recently discussed in the media, while the request with response categories provides a frame of reference indicating what is expected from the respondent. The option of "other" category along with a set of responses can be introduced but it turns out that this option is not chosen as frequently as expected. Hence, the authors concluded that the given response categories of a request guide respondents in their answer choices.

Subsequent research by Krosnick and Schuman (1988) suggests that there is more consistency across the open and closed request results if the coding of the answers of the open request is more in line with the categories used by less-educated people. This brought Krosnick and Fabrigar (1997) to conclude that open requests are preferable because the effect of the researcher on the result is avoided.

The last statement may be correct for the abovementioned type of request, where a choice out of a multitude of nominal categories is requested, however, the findings need to be investigated further to determine whether they are also true for other open requests for an answer. Therefore, let us explore some other possibilities.

Krosnick and Fabrigar (forthcoming) indicate in another chapter of their book that not all open requests can be trusted at face value. They discuss the open "WHY request and the validity of introspection." In psychology, introspection has been discussed at length by the different schools of thought where some scholars think that only people can know why they do things, and therefore they should be asked. Other scholars argue that answers based on introspection cannot be trusted. One of the reasons provided is quick memory loss of thoughts concerning the choices made. Therefore a "think aloud" procedure is suggested, but if one asks for arguments before or while people are making choices, this in itself can influence the process (Ericson and Simon 1984; Wilson and Dunn 1986) and most of the time rationalizations of the answer choice are provided. This is not only the view of the behaviorists like Skinner (1953), but also of scholars with a less extreme point of view (Nisbett and Wilson, 1977).

Krosnick and Fabrigar (1997), while applying this bulk of research on survey research, comment "...if results based on introspection requests seem sensible on their surface, we would all be inclined to view them as valid. And yet, as Nisbett and Wilson (1977) and Wilson and Dunn (1986) have made clear, this

apparent sensibility may well be the result of people's desire to appear rational, rather than the result of actual validity of introspection." Therefore Krosnick and Fabrigar (1997) clearly indicate their reservations with the use of introspection procedure with the open request for an answer method. One should, however, also remark that formulating alternative procedures for introspection is not very easily done.

Wouters (2001), in her research, has specified open requests for all kinds of combinations of concepts and request forms that have been mentioned in Chapters 2, 3, and 4. For example one could ask:

5.2 *How would you evaluate the presidency of Clinton?*

It is clear that an evaluation is asked but the possible responses are not specified. So, this is an open request, and the respondent can give an answer in many different ways. In a similar way Wouters (2001) was able to transform nearly all possible closed requests into open-ended requests for answers. Hence the pertinent question is which of the two forms is better. To answer this question, a lot of research is still needed. Presently, we can say only that closed requests are more efficient than open requests because the former do not require an extra coding phase.

The analysis of Wouters (2001) also showed that it is not always simple to formulate a closed form for all open requests. We will demonstrate our point with the following example:

5.3a *What do you think about the presidency of Clinton?*

Example 5.3a is an open-ended request, however what is special about this request is that it does not measure a specific concept because respondents can answer with an evaluation (good or bad) but also with a cognition (that Clinton's government was the first to balance the budget) or a relationship (that Clinton's presidency led to an impeachment procedure) as just a few examples of possible answers. Not only is the answer open-ended but also the concept itself that is measured. Our hypothesis is that such requests are used to determine what aspect of the object the respondents consider the most important from which are derived further requests about this aspect. If that is true, an alternative in closed form to the open-ended request could be

5.3b *What is for you the most important aspect of the presidency of Clinton?*

1. *His foreign policy*
2. *His national policy*
3. *His economic policies*
4. *His personal conduct*
5. *Others*

Another type of open request that is hard to formulate in closed form concerns the enumeration of different events of actions. An example is

5-4 *Can you describe the different events that took place before the demonstration changed into a violent action?*

Here the respondent has to provide a series of events that have occurred in sequence.

From example 5-4 it can be inferred that asking an equivalent request in closed format would require a very different and complex series of requests.

Another type of request for an answer that requires special attention is a request for a frequency or an amount. Examples are found below:

5-5a *How many hours did you watch TV last night?*

5-5b *How much did you pay for your car?*

These requests are in some sense the opposite of the open requests we have discussed above, because now it is very clear how the respondents have to answer. The first request asks for a number that indicates the number of hours they have watched TV, and the second asks for a monetary amount. So people know quite well how they should answer, but nevertheless the answer is open because no response options have been provided to them (Tourangeau et al. 2000). For these requests that ask numeric answers, closed alternatives have been formulated. They will be discussed in further detail in the section on vague quantifiers.

It might depend on what request type we are about to use whether we choose an open or closed form. For most open requests alternatives in closed form exist; for others, alternative closed requests are difficult to formulate. For those requests that can be asked in a variety of ways, different aspects should be considered. First, it is important to consider whether more information is obtained through using the open request format. If that is not the case, then it is better to choose the closed form because the processing of the information is much easier. A second issue is, whether open and closed requests lead to different response distributions and relationships with other variables. If that is the case, one has to consider which request form is better. Evaluation of the effects on the data quality will be discussed later. If the same results are obtained or the quality is not clearly better for the open requests, then the closed requests should be preferred because of the efficiency in information processing. It will be clear that in our opinion the conclusion of Krosnick and Fabrigar (1997) is still premature and we think that further research is required before a conclusion about the choice between open and closed requests can be stated with certainty. We speculate that the request choice will depend on the type of issue the request is aiming at as was the case with our examples.

5-2 CLOSED CATEGORICAL REQUESTS

The first of the requirements regarding closed response answer categories is that they should be *complete*. In practice, however, sometimes the answer alternatives are not complete, which can result in nonresponse. Such an example is given below:

5-6a *What is the composition of your household?*

1. One single adult
2. Two adults
3. Two adults and one child
4. Two adults with two children
5. Two adults with three children
6. One adult with one child

After scanning the answer options for 5-6a, it becomes clear that the answer categories are not exhaustive since there are several variations of adults and children possible and one for communes is missing. Hence 5-6b is a more complete version:

5-6b *What is the composition of your household?*

1. Number of adults ...
2. Number of children ...

The second requirement is that the answer categories are *exclusive*, or in other words they should not overlap. An example of overlapping answer categories is found in request 5-7a:

5-7a *What is the most important reason why you are against nuclear energy?*

1. Too expensive
2. Too dangerous
3. Causes environmental problems
4. Other

In request 5-7a the second and third categories are not exclusive because environmental problems can cause dangers and dangers, like radioactive waste, can cause environmental problems. Therefore, a respondent may be confused about which choice to make. The remedy is to reformulate these two categories in order to make them exclusive:

5-7b *What is the most important reason why you are against nuclear energy?*

1. Too expensive
2. The probability of an accident is too high
3. Too much radioactive waste
4. Other ...

Here the second category focuses on accidents and the third, on radioactive waste, which are now distinct and no longer overlap.

A third requirement is that answer categories *match* with the information provided in the request or statement asked (Lessler and Forsight 1996; Graesser et al. 2000a,b):

5.8a How far do you agree or disagree with the statement that governmental decisions are always carried out

1. Completely agree
2. Agree
3. Neither agree nor disagree
4. Disagree

In the example the statement refers to an objective concept (a behavior), while the answer categories relate to subjective concepts. The appropriate answer categories would be “true/false.” The request could be reformulated in the following manner:

5.8b Do you think that the following statement is true or false?
Governmental decisions are always carried out.

1. True
2. Neither true nor false
3. False

Finally, a requirement is that all the response categories represent *the same concept*. Sometimes a mismatch of answer categories occurs because they concern different concepts and then it is difficult for the respondent to choose a category. Example 5.9 illustrates a case where this is not correct:

5.9 What is your opinion about a ban on driving a car in downtown areas?

1. Inconvenient
2. Acceptable

The first category refers to a feeling, while the second is a right. In order to be consistent, it is possible to provide either a feeling (unpleasant/pleasant) or a right (acceptable/unacceptable) as options of the uncertainty space. All requests for an answer with closed answer categories should satisfy the abovementioned requirements.

In the following sections we want to illustrate the different types of response categories that are available to the survey designer. The first type uses nominal categories without any ordering, while the second type provides ordinal response categories and the third consists of what is called vague quantifiers.

5.2.1 Nominal categories

Requests for an answer using unordered response categories are an alternative for the open requests asking for one option out of a set. An example is

5.10 What is the most important problem that our country faces at the moment?

1. Terrorism
2. Unemployment
3. Racism

4. Criminality
5. Other, please specify ...

Similar requests can be asked for the most important aspect of the work and many other topics. There is no ordering in the different response possibilities even though they can be numbered in the questionnaire and certainly in the database, but, the numbers cannot suggest an ordering on any dimension because that dimension does not exist. Response scales that are not ordered are called *nominal scales*.

A special nominal scale is a scale for *dichotomous* responses where only two answers are possible for example:

5.11 Did you vote in the last elections?
1. No
2. Yes

In this case the scale is officially nominal, indicating no ordering. However, it is possible to use the scale in the ordinal sense and apply analyses that at minimum require ordinal data and it is arbitrary if the coding by the researcher is completed as 0–1 or 1–2 for the dichotomous scale.

5.2.2 Ordinal scales

Ordinal response categories require that there is an ordering of the response categories. Such sets of response alternatives are very common in subjective judgments. For example

5.12 How good do you think Clinton was as president?
1. Very bad
2. Bad
3. Neither good nor bad
4. Good
5. Very good

In this case there is an ordering in the response categories, and one can say that the numbers in front of the categories suggest an ordered scale where 1 is the lowest and 5 is the highest category. Similar scales can be made with any predicate with “high” and “low,” “friendly” and “unfriendly,” “active” and “passive,” to name only a few examples.

Although such an ordinal scale is called a 5-point scale – a scale with 5 possible answers – a person with a positive evaluation of Clinton has only two possibilities: good or very good. If it is desirable to have a more precise answer, it can be specified as a 7-point scale such as the one below:

1. Very bad
2. Rather bad
3. Bad
4. Neither good nor bad
5. Good

6. *Rather good*
7. *Very good*

Along the same line one can also construct a 9- or 11-point scale. Keep in mind that there is a limit to the possibilities of labels for the different categories, and that it is also possible to specify ordinal scales with labels for only a limited number of categories. Common examples are the following:

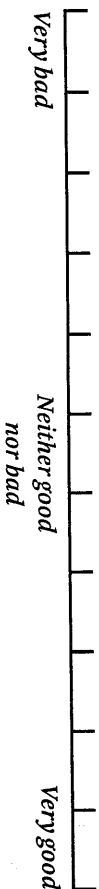
5-13a *How good do you think Clinton was as president?*

*Express your opinion in a number between 0 and 10, where
0 = very bad and 10 = very good*

or

5-13b *How good do you think Clinton was as president?*

Express your opinion by placing an x at the point of the scale that expresses your opinion the best



Examples 5-13a and 5-13b are both 11-point scales, the distinction is that the former has only two labeled categories while the latter has three labeled categories; and that the first request uses numbers while the second is a typical example of what is called a *rating scale*.

Many alternative presentations can be developed with ordinal response scales. What is important is that the categories are ordered in some way from low to high. It can also be done by pictures of faces that are more or less happy or ladders where each step indicates a different level of satisfaction (Andrews and Withey 1974) or a thermometer where the increasing grades indicate the warmth of the feelings of respondents toward parties and party leaders. The United States' National Election Studies are exemplar for this type of creative ordinal response scale grading.

When developing ordinal scales a range of decisions is at the researchers' disposal. We will discuss some of these choices with their alternatives. First, we have seen that either all or some of the possible responses can be labeled. Therefore, the responses can be *completely labeled* or *partly labeled*.

In example 5-13a the numbers in front of the categories were ordered in the same way from low to high as the labels and they started with the lowest or most negative category. It can also happen that there is no *correspondence* between the category labels and the numbers or that the scale does not go from low or negative to positive but vice versa.

All the scales presented so far are *symmetric* around the middle of the scale, which means that there are as many categories at the positive as at the negative side.

In general it is advisable to use symmetric scales; the reason can be demonstrated by the example:

- 5-14
1. *Very unhappy*
 2. *Unhappy*
 3. *Neither unhappy nor happy*
 4. *Happy*
 5. *Rather happy*
 6. *Very happy*

This example demonstrates that it appears awkward to be using an asymmetric scale in this case. However, if we know that all respondents' answers are on the happy side of the scale, it is not very efficient to use a 5-point scale from "very unhappy" to "very happy" because the distribution of happiness in the population is reduced to a 2-point scale. Therefore, an asymmetric 5-point scale is more appropriate and precise:

- 5-15
1. *Not happy*
 2. *A bit happy*
 3. *Happy*
 4. *Rather happy*
 5. *Very happy*

Example 5-15 has a 5-point scale that favors the positive side, while the "not happy" side of the scale, is represented by only one response category. Such a scale presupposes knowledge about the happiness of a survey population; otherwise, such an asymmetric scale is biased.

So far, except for in the last example, all sets of response scales were also *bipolar*, which means that there are two opposite sides of the scales: positive to negative or active to passive. The last scale of happiness was made one-sided or *unipolar*, but happiness itself is in principle a bipolar concept, going from unhappy to happy. Therefore we also said that the unipolar scale presupposed knowledge of the distribution of feelings within the population. There are, however, also concepts that are typically unipolar. For example, "attachment to a party" goes from "no attachment" to "strong attachment" because it is impossible to imagine a negative side of the scale of attachment.

The discussion above has served to demonstrate that both the provided scale for responses and the concept can be in *agreement* with each other (both bipolar or both unipolar) or in *disagreement* if the concept is bipolar, but the responses are only unipolar, as in example 5-15.

So far we have used a *neutral category* or a *middle category*, but it is not always necessary to do so. If it is necessary to force people to make a choice in a specific direction, then the middle category can be omitted. Schuman and Presser (1981) have shown that this has no effect on the distribution of the respondents over the positive and negative categories. However, it might have the effect that fewer people are willing to answer the request because, according to them, their response is not provided and consequently they choose for a "don't know" or "refusal" (Klingemann 1997).

The “*don’t know*” category has been the subject of serious investigation. Research has centered around the question of whether it should be offered, and if so, in what form. One can ask, for instance, before the request itself is asked whether people have an opinion or not about the topic in question. This is the most explicit “*don’t know*” check. The second possibility is to provide “*don’t know*” explicitly as one of the response options. The third possibility is that “*don’t know*” is not mentioned but that it is an admissible response alternative that can be found on the questionnaire of the interviewer but is not mentioned as a possibility to the respondent. Finally, there is the possibility of omitting it altogether.

Providing the “*don’t know*” option explicitly creates several obstacles. The most important issue is that respondents can choose this option for several reasons which have nothing to do with their own opinion. Krosnick and Fabrigar (forthcoming) mention that this option is chosen because respondents don’t want more requests or because they do not want to think about the request and therefore an acceptable option “*don’t know*” is easily available. The authors call this “satisficing behavior of a respondent.”

Schuman and Presser (1981) argue that people who normally would say that they “*don’t know*” would make a difference in the relationships between variables under investigation. They report on a study where without respondents using the “*don’t know*” category the correlation between two variables was close to zero while with them it went up to .6.

Another problem with people choosing “*don’t know*” is that fewer representatives of the population are left for the analysis. If the option is available for several requests, the number of people with complete data on a larger set of variables can decrease and it becomes questionable whether the respondents who are left in the sample are on the whole representative for the population. These three arguments have led researchers to allow for the “*don’t know*” option, but only if the respondent explicitly asks for it. However, whether this is the most scientific course of action, we will evaluate later.

So far the focus of our discussion has been specification of response categories for subjective variables. However, ordinal response categories are also used for objective variables such as the frequency of activities or categories of income and prices. An example could be

5.16a How often do you watch TV during a week?

1. Very often
2. Often
3. Regularly
4. Seldom
5. Never

If we had omitted the response alternatives, this could have been an open-ended request, but researchers often add response categories to such requests and the issue is that respondents can differ in their interpretation of the

different labels: what is “*often*” for one person means “*seldom*” for another. It all depends on the reference point of the respondent. Therefore these ordinal scales are called *vague quantifiers*. We could have also asked the following:

5.16b How often do you watch TV during a week?

1. Every day
2. 5 or 6 times
3. 3 or 4 times a week
4. 1 or 2 times a week
5. Never

This request is more precise and less prone to different interpretations. Even so, 5.16b is an ordinal scale because it is not clear what numeric values the categories 2–4 represent.

Table 5.1: The results of Hippler and Schwartz with respect to TV watching

Categories at the low side	Percentage of respondents	Categories at the high side	Percentage of respondents
< ½ hour	11.5		
½ - 1 ½ hours	53.8		
1 ½ - 2 ½ hours	34.7	< 2 ½ hours	70.6
> 2 ½ hours	0.0	> 2 ½ hours	29.4
Total	100	Total	100

Similar scales can be used for income and prices with the option of using vague quantifiers or more precise category labels. Hippler and Schwarz (1987) made a remarkable observation when they varied the category labels in an experiment about the amount of time people watch TV. In it they did not use vague quantifiers like those of example 5.16a but two different and separate categories for the number of TV viewing hours. Their results are presented in Table 5.1. The table shows that the different categories had a considerable effect on the responses. Their explanation was that respondents do not have an answer readily available for this type of request. Instead, they use the response scale as their frame of reference. Respondents estimate their TV watching time on whether they view themselves as more or less TV watching than other persons. Therefore, if they consider that they watch more TV than others, they will choose the high end of the scale and vice versa. This experiment shows that even for objective variables the answers do not represent absolute judgments but relative judgments. It has been suggested that people always make relative judgments. If that is so, it is better to adjust the approach of asking requests to the human judgment factor. We will investigate this problem in the next section in more depth.

5.2.3 Continuous scales

Another form for response possibilities is to give respondents instructions to express their opinions in numbers or lines. This approach was developed with the idea that it would result in more precise information than would the other methods discussed in previous sections.

Originally such approaches were used in psychophysics. For an overview we refer to Stevens (1975). Presently these measurement devices have been introduced in the social sciences by Hamblin (1974), Saris et al. (1977), Wegener (1982) and Lodge (1981). When these approaches were introduced special procedures, called *magnitude estimation* and *line production*, were used. The basic idea is rather simple and will be illustrated by several examples of procedures used in practice. Originally a request for an answer was formulated as follows:

5.17a Occupations differ with respect to status. We would like to ask you to estimate the status of a series of occupations. If we give the status of the occupation of a schoolteacher a score of 100, how would you evaluate the other occupations? If an occupation has a status that is twice as high as that of a schoolteacher, give a twofold larger number or 200. If the status of the occupation is half that of a schoolteacher, divide by 2, which gives 50.

What is the status of a physician? ...

Of a carpenter? ...

And so on.

People are asked to match the ratios of status judgements with the ratios of numbers. This could also be done using "line production," as has been shown in the following instruction:

5.17b Occupations differ with respect to status. We would like to ask you to estimate the status of a series of occupations. We express the status of the occupation of a schoolteacher by a standard line as follows:

If an occupation has a status that is twice as high as that of a schoolteacher, draw a line which is twice as long. If the status of the occupation is half of that of a schoolteacher, draw a line that is half the size of the standard line.
What is the status of a physician?

Of a carpenter?

With these procedures a striking precision of responses was obtained (Hamblin 1974; Saris et al. 1977). However, in their embryonic stage these approaches were used only for evaluation of stimuli, as we have previously indicated. Currently other concepts are also measured in this way. For example, we could reformulate the frequently asked satisfaction request using continuous scales as follows:

5.18 a How satisfied are you with your house? Express your opinion with a number between 0 and 100, where 0 is completely dissatisfied and 100 is completely satisfied.

This request differs in several points from the original instruction. The first point is that the ratio estimation is no longer mentioned. The reason is that the results are not very different whether one gives this instruction explicitly, while at the same time, omitting this instruction makes the formulation much simpler. The second point is that two *reference points* have been mentioned instead of just one. This is due to research showing that people use different scales to answer these requests if only one reference point is provided, while using two reference points it is less of a concern (Saris 1988b). A condition for this conclusion is that *fixed reference points* are used. With fixed reference points, we mean that there is no doubt about the position of the reference point on the subjective scale in the mind of the respondent. For example, "completely dissatisfied" and "completely satisfied" must be the endpoints of the opinion scale of the respondent. If we would use "dissatisfied" and "satisfied" as reference points, then respondents may vary in their interpretation of these terms because some of them see them as endpoints of the scales while others do not.

The disadvantage of using numbers is that people tend to use numbers which can be divided by 5 (Tourangeau et al. 2000). This leads to rather peaked distributions of the results. This can be largely avoided by the use of line length instead of requesting a numerical evaluation. For request 5.18a the instruction, using line length as response mode, would be as follows:

5.18b How satisfied are you with your house? Express your opinion in length of lines, where completely dissatisfied is expressed by the following line:

and completely satisfied by the following line:

Now express your opinion by drawing a line representative of your opinion:

The disadvantage of the line production is, of course, that later the lines need to be measured. This is a challenge if paper-and-pencil procedures for data collection are used but with computer assisted interviewing (CAI) the programs can measure the length of lines routinely.

