Figure 8.4 Enriching a focus group with personas displayed on the wall for all participants to see

As another example, Clara Mancini et al. (2009) used a combination of questionnaire prompts and deferred contextual interviews when investigating mobile privacy. A simple multiple-choice questionnaire was sent electronically to the participants' smartphones, and they answered the questions using these devices. Interviews about the recorded events were conducted later, based on the questionnaire answers given at the time of the event.

8.5   Questionnaires

Questionnaires are a well-established technique for collecting demographic data and users' opinions. They are similar to interviews in that they can have closed or open-ended questions, but once a questionnaire is produced, it can be distributed to a large number of participants without requiring additional data gathering resources. Thus, more data can be collected than would normally be possible in an interview study. Furthermore, participants who are located in remote locations or those who cannot attend an interview at a particular time can be involved more easily. Often a message is sent electronically to potential participants directing them to an online questionnaire.

Effort and skill are needed to ensure that questions are clearly worded and the data collected can be analyzed efficiently. Well-designed questionnaires are good for getting answers to specific questions from a large group of people. Questionnaires can be used on their own

or in conjunction with other methods to clarify or deepen understanding. For example, information obtained through interviews with a small selection of interviewees might be corroborated by sending a questionnaire to a wider group to confirm the conclusions.

Questionnaire questions and structured interview questions are similar, so which technique is used when? Essentially, the difference lies in the motivation of the respondent to answer the questions. If their motivation is high enough to complete a questionnaire without anyone else present, then a questionnaire will be appropriate. On the other hand, if the respondents need some persuasion to answer the questions, a structured interview format would be better. For example, structured interviews are easier and quicker to conduct if people will not stop to complete a questionnaire, such as at a train station or while walking to their next meeting.

It can be harder to develop good questionnaire questions compared with structured interview questions because the interviewer is not available to explain them or to clarify any ambiguities. Because of this, it is important that questions are specific; when possible, ask closed-ended questions and offer a range of answers, including a "no opinion" or "none of these" option. Finally, use negative questions carefully, as they can be confusing and may lead to false information. Some questionnaire designers, however, use a mixture of negative and positive questions deliberately because it helps to check the users' intentions.

8.5.1 Questionnaire Structure

Many questionnaires start by asking for basic demographic information (gender, age, place of birth) and details of relevant experience (the number of hours a day spent searching on the Internet, the level of expertise within the domain under study, and so on). This background information is useful for putting the questionnaire responses into context. For example, if two responses conflict, these different perspectives may be because of their level of experience—a group of people who are using a social networking site for the first time are likely to express different opinions than another group with five years' experience of using such sites. However, only contextual information that is relevant to the study goal needs to be collected. For example, it is unlikely that a person's height will provide relevant context to their responses about Internet use, but it might be relevant for a study concerning wearables.

Specific questions that contribute to the data-gathering goal usually follow these demographic questions. If the questionnaire is long, the questions may be subdivided into related topics to make it easier and more logical to complete.

The following is a checklist of general advice for designing a questionnaire:

- Think about the ordering of questions. The impact of a question can be influenced by question order.
- Consider whether different versions of the questionnaire are needed for different populations.
- Provide clear instructions on how to complete the questionnaire, for example, whether answers can be saved and completed later. Aim for both careful wording and good typography.
- Think about the length of the questionnaire, and avoid questions that don't address the study goals.
- If the questionnaire has to be long, consider allowing respondents to opt out at different stages. It is usually better to get answers to some sections than no answers at all because of dropout.
- Think about questionnaire layout and pacing; for instance, strike a balance between using white space, or individual web pages, and the need to keep the questionnaire as compact as possible.

8.5.2   Question and Response Format

Different formats of question and response can be chosen. For example, with a closed-ended question, it may be appropriate to indicate only one response, or it may be appropriate to indicate several. Sometimes, it is better to ask users to locate their answer within a range. Selecting the most appropriate question and response format makes it easier for respondents to answer clearly. Some commonly used formats are described next.

Check Boxes and Ranges

The range of answers to demographic questions is predictable. Nationality, for example, has a finite number of alternatives, and asking respondents to choose a response from a predefined list makes sense for collecting this information. A similar approach can be adopted if details of age are needed. But since some people do not like to give their exact age, many questionnaires ask respondents to specify their age as a range. A common design error arises when the ranges overlap. For example, specifying two ranges as 15–20 and 20–25 will cause confusion; that is, which box do people who are 20 years old check? Making the ranges 15–19 and 20–24 avoids this problem.

A frequently asked question about ranges is whether the interval must be equal in all cases. The answer is no—it depends on what you want to know. For example, people who might use a website about life insurance are likely to be employed individuals who are 21–65 years old. The question could, therefore, have just three ranges: under 21, 21–65, and over 65. In contrast, to see how the population's political views vary across generations might require 10-year cohort groups for people over 21, in which case the following ranges would be appropriate: under 21, 21–30, 31–40, and so forth.

Rating Scales

There are a number of different types of rating scales, each with its own purpose (see Oppenheim, 2000). Two commonly used scales are the Likert and semantic differential scales. Their purpose is to elicit a range of responses to a question that can be compared across respondents. They are good for getting people to make judgments, such as how easy, how usable, and the like.

Likert scales rely on identifying a set of statements representing a range of possible opinions, while semantic differential scales rely on choosing pairs of words that represent the range of possible opinions. Likert scales are more commonly used because identifying suitable statements that respondents will understand consistently is easier than identifying semantic pairs that respondents interpret as intended.

Likert Scales

Likert scales are used for measuring opinions, attitudes, and beliefs, and consequently they are widely used for evaluating user satisfaction with products. For example, users' opinions about the use of color in a website could be evaluated with a Likert scale using a range of numbers, as in question 1 here, or with words as in question 2:

1. The use of color is excellent (where 1 represents strongly agree and 5 represents strongly disagree):

1	2	3	4	5
☐	☐	☐	☐	☐

2. The use of color is excellent:

Strongly agree	Agree	OK	Disagree	Strongly disagree
☐	☐	☐	☐	☐

In both cases, respondents would be asked to choose the right box, number, or phrase. Designing a Likert scale involves the following steps:

1. *Gather a pool of short statements about the subject to be investigated.* Examples are "This control panel is clear" and "The procedure for checking credit rating is too complex." A brainstorming session with peers is a good way to identify key aspects to be investigated.
2. *Decide on the scale.* There are three main issues to be addressed here: How many points does the scale need? Should the scale be discrete or continuous? How can the scale be represented? See Box 8.4 What Scales to Use: Three, Five, Seven, or More? for more on this.
3. *Select items for the final questionnaire, and reword as necessary to make them clear.*

In the first example above, the scale is arranged with 1 as the highest choice on the left and 5 as the lowest choice on the right. The logic for this is that first is the best place to be in a race and fifth would be the worst place. While there is no absolute right or wrong way of ordering the numbers other researchers prefer to arrange the scales the other way around with 1 as the lowest on the left and 5 as the highest on the right. They argue that intuitively the highest number suggests the best choice and the lowest number suggests the worst choice. Another reason for going from lowest to highest is that when the results are reported, it is more intuitive for readers to see high numbers representing the best choices. The important thing is to be consistent.

Semantic Differential Scales

Semantic differential scales explore a range of bipolar attitudes about a particular item, each of which is represented as a pair of adjectives. The participant is asked to choose a point between the two extremes to indicate agreement with the poles, as shown in Figure 8.5. The score for the investigation is found by summing the scores for each bipolar pair. Scores are then computed across groups of participants. Notice that in this example the poles are mixed so that good and bad features are distributed on the right and the left. In this example, there are seven positions on the scale.
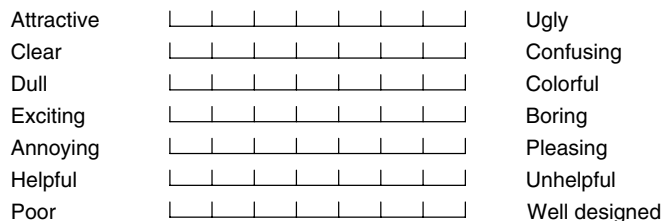
Attractive	|__|__|__|__|__|__|	Ugly
Clear	|__|__|__|__|__|__|	Confusing
Dull	|__|__|__|__|__|__|	Colorful
Exciting	|__|__|__|__|__|__|	Boring
Annoying	|__|__|__|__|__|__|	Pleasing
Helpful	|__|__|__|__|__|__|	Unhelpful
Poor	|__|__|__|__|__|__|	Well designed

Figure 8.5 An example of a semantic differential scale

BOX 8.4

What Scales to Use: Three, Five, Seven, or More?

Issues to address when designing Likert and semantic differential scales include the following: how many points are needed on the scale, how should they be presented, and in what form?

Many questionnaires use seven- or five-point scales, and there are also three-point scales. Some even use nine-point scales. Arguments for the number of points go both ways. Advocates of long scales argue that they help to show discrimination. Rating features on an interface is more difficult for most people than, say, selecting among different flavors of ice cream, and when the task is difficult, there is evidence to show that people "hedge their bets." Rather than selecting the poles of the scales if there is no right or wrong, respondents tend to select values nearer the center. The counterargument is that people cannot be expected to discern accurately among points on a large scale, so any scale of more than five points is unnecessarily difficult to use.

Another aspect to consider is whether to give the scale an even or odd number of points. An odd number provides a clear central point, while an even number forces participants to decide and prevents them from sitting on the fence.

We suggest the following guidelines:

How many points on the scale?
Use a small number, three, for example, when the possibilities are very limited, as in Yes/No type answers.

☐   ☐   ☐
Yes   Don't know   No

Use a medium-sized range, five, for example, when making judgments that involve like/dislike or agree/disagree statements.

Strongly agree	Agree	OK	Disagree	Strongly disagree
☐	☐	☐	☐	☐

Use a longer range, seven or nine, for example, when asking respondents to make subtle judgments, such as when asking about a user experience dimension such as "level of appeal" of a character in a video game.

very appealing   ok   repulsive

Discrete or continuous?
Use boxes for discrete choices and scales for finer judgments.

What order?
Decide which way to order the scale, and be consistent. ■

ACTIVITY 8.3

Spot four poorly designed features in the excerpt from a questionnaire in Figure 8.6.

Figure 8.6 A questionnaire with poorly designed features

Comment

Some of the features that could be improved upon include the following:

- Question 2 requests an exact age. Many people prefer not to give this information and would rather position themselves within a range.
- In question 3, the number of hours spent searching is indicated with overlapping scales, that is, 1–3 and 3–5. How would someone answer if they spend 3 hours a day searching online?
- For question 4, the questionnaire doesn't say how many boxes to check.
- The space left for people to answer open-ended question 5 is too small, which will annoy some people and deter them from giving their opinions.

Many online survey tools prevent users from making some of these design errors. It is important, however, to be aware of such things because paper is still sometimes used. ■

8.5.3 Administering Questionnaires

Two important issues when using questionnaires are reaching a representative sample of participants and ensuring a reasonable response rate. For large surveys, potential respondents need to be selected using a sampling technique. However, interaction designers commonly use a small number of participants, often fewer than 20 users. Completion rates of 100 percent are often achieved with these small samples, but with larger or more remote populations, ensuring that surveys are returned is a well-known problem. A 40 percent return is generally acceptable

for many surveys, but much lower rates are common. Depending on your audience, you might want to consider offering incentives (see section 8.2.3, "Relationship with Participants").

While questionnaires are often online, paper questionnaires may be more convenient in some situations, for example, if participants do not have Internet access or if it is expensive to use. Occasionally, short questionnaires are sent within the body of an email, but more often the advantages of the data being compiled automatically and either partly or fully analyzed make online questionnaires attractive. Online questionnaires are interactive and can include check boxes, radio buttons, pull-down and pop-up menus, help screens, graphics, or videos (see Figure 8.7). They can also provide immediate data validation; for example, the entry must be a number between 1 and 20, and automatically skip questions that are irrelevant to some respondents, such as questions aimed only at teenagers. Other advantages of online questionnaires include faster response rates and automatic transfer of responses into a database for analysis (Toepoel, 2016).

Figure 8.7 An excerpt from a web-based questionnaire showing check boxes, radio buttons, and pull-down menus

The main problem with online questionnaires is the difficulty of obtaining a random sample of respondents; online questionnaires usually rely on convenience sampling, and hence their results cannot be generalized. In some countries, online questions, often delivered via smartphones, are frequently used in conjunction with television to elicit viewers' opinions of programs and political events.

Deploying an online questionnaire involves the following steps (Toepoel, 2016, Chapter 10):

1. *Plan the survey timeline*. If there is a deadline, work backward from the deadline and plan what needs to be done on a weekly basis.
2. *Design the questionnaire offline*. Using plain text is useful as this can then be copied more easily into the online survey tool.
3. *Program the online survey*. How long this will take depends on the complexity of the design, for example, how many navigational paths it contains or if it has a lot of interactive features.
4. *Test the survey, both to make sure that it behaves as envisioned and to check the questions themselves*. This includes getting feedback from content experts, survey experts, and potential respondents. This last group forms the basis of a pilot study.
5. *Recruit respondents*. As mentioned earlier, participants may have different reasons for taking part in the survey, but especially when respondents need to be encouraged, make the invitations intriguing, simple, friendly, respectful, trustworthy, motivating, interesting, informative, and short.

There are many online questionnaire templates available that provide a range of options, including different question types (for example open-ended, multiple choice), rating scales (such as Likert, semantic differential), and answer types (for example, radio buttons, check boxes, drop-down menus).

The following activity asks you to make use of one of these templates. Apart from being able to administer an online questionnaire widely, these templates also enable the questionnaire to be segmented. For example, airline satisfaction questionnaires often have different sections for check-in, baggage handling, airport lounge, inflight movies, inflight food service, and so forth. If you didn't use an airport lounge or check your baggage, you can skip those sections. This avoids respondents getting frustrated by having to go through questions that are not relevant to them. It is also a useful technique for long questionnaires, as it ensures that if a respondent opts out for lack of time or gets tired of answering the questions, the data that has been provided already is available to be analyzed.

ACTIVITY 8.4

Go to questionpro.com, surveymonkey.com, or a similar survey site and design your own questionnaire using the set of widgets that is available for a free trial period.

Create an online questionnaire for the set of questions that you developed for Activity 8.2. For each question, produce two different designs; for example, use radio buttons and drop-down menus for one question, and provide a 10-point semantic differential scale and a 5-point scale for another question.

What differences (if any) do you think the two designs will have on a respondent's behavior? Ask a number of people to answer one or the other of your questions and see whether the answers differ for the two designs.

Comment
Respondents may have used the response types in different ways. For example, they may select the end options more often from a drop-down menu than from a list of options that are chosen via radio buttons. Alternatively, you may find no difference and that people's opinions are not affected by the widget style used. Some differences, of course, may be due to the variation between individual responses rather than being caused by features in the questionnaire design. To tease the effects apart, you would need to ask a large number of participants (for instance, in the range 50–100) to respond to the questions for each design. ■

BOX 8.5

Do people answer online questionnaires differently than paper and pencil? If so, why?

There has been much research examining how people respond to surveys when using a computer compared with paper and pencil methods. Some studies suggest that people are more revealing and consistent in their responses when using a computer to report their habits and behaviors, such as eating, drinking, and amount of exercise see Luce et al. (2003). Students have also been found to rate their instructors less favorably when online (Chang, 2004).

In a Danish study in which 3,600 people were invited to participate, the researchers concluded that although response rates for web-based invitations were lower, they were more cost-effective (by a factor of 10) and had only slightly lower numbers of missing values than questionnaires sent via paper (Ebert et al., 2018). Similarly, a study by Diaz de Rada and Dominguez-Alvarez (2014), in which the quality of the information collected from a survey given to citizens of Andalusia in Spain was analyzed, several advantages of using online versus paper-based questionnaires were identified. These included a low number of unanswered questions, more detailed answers to open-ended questions, and longer answers to questions in the online questionnaires than in the paper questionnaires. In the five open-ended questions, respondents wrote 63 characters more on average on the online questionnaires than on the paper questionnaires. For the questions in which participants had to select from a drop-down menu, there was a better response rate than when the selection was presented on paper with blank spaces.

One factor that can influence how people answer questions is the way the information is structured, such as the use of headers, the ordering, and the placement of questions. Online questionnaires provide more options for presenting information, including the use of drop-down menus, radio buttons, and jump-to options, which may influence how people read and navigate a questionnaire. But do these issues affect respondents' answers? Smyth et al. (2005) have found that providing forced choice formats results in more options being selected. Another example is provided by Funcke et al. (2011), who found that continuous sliders enabled researchers to collect more accurate data because they support continuous rather than discrete scales. They also encouraged higher response rates. What can be concluded from these investigations is that the details of questionnaire design can impact how respondents react. ■