Chapter 13

Asking users and experts

13.1 Introduction

In the last chapter we looked at observing users. Another way of finding out what users do, what they want to do, like, or don't like is to ask them. Interviews and questionnaires are well-established techniques in social science research, market research, and human-computer interaction. They are used in "quick and dirty" evaluation, in usability testing, and in field studies to ask about facts, behavior, beliefs, and attitudes. Interviews and questionnaires can be structured (as in the HutchWorld case study in Chapter 10), or flexible and more like a discussion, as in field studies. Often interviews and observation go together in field studies, but in this chapter we focus specifically on interviewing techniques.

The first part of this chapter discusses interviews and questionnaires. As with observation, these techniques can be used in the requirements activity (as we described in Chapter 7), but in this chapter we focus on their use in evaluation. Another way of finding out how well a system is designed is by asking experts for their opinions. In the second part of the chapter, we look at the techniques of heuristic evaluation and cognitive walkthrough. These methods involve predicting how usable interfaces are (or are not). As in the previous chapter, we draw on the DECIDE framework from Chapter 11 to help structure studies that use these techniques.

The main aims of this chapter are to:

- Discuss when it is appropriate to use different types of interviews and questionnaires.
- Teach you the basics of questionnaire design.
- Describe how to do interviews, heuristic evaluation, and walkthroughs.
- Describe how to collect, analyze, and present data collected by the techniques mentioned above.
- Enable you to discuss the strengths and limitations of the techniques and select appropriate ones for your own use.

13.2 Asking users: interviews

Interviews can be thought of as a "conversation with a purpose" (Kahn and Cannell, 1957). How like an ordinary conversation the interview is depends on the questions to be answered and the type of interview method used. There are four main types of interviews: *open-ended or unstructured, structured, semi-structured,* and *group* interviews (Fontana and Frey, 1994). The first three types are named according to how much control the interviewer imposes on the conversation by following a *predetermined set of questions*. The fourth involves a small group guided by an interviewer who facilitates discussion of a specified set of topics.

The most appropriate approach to interviewing depends on the evaluation goals, the questions to be addressed, and the paradigm adopted. For example, if the goal is to gain first impressions about how users react to a new design idea, such as an interactive sign, then an informal, open-ended interview is often the best approach. But if the goal is to get feedback about a particular design feature, such as the layout of a new web browser, then a structured interview or questionnaire is often better. This is because the goals and questions are more specific in the latter case.

13.2.1 Developing questions and planning an interview

When developing interview questions, plan to keep them short, straightforward and avoid asking too many. Here are some guidelines (Robson, 1993):

Avoid long questions because they are difficult to remember.

- Avoid compound sentences by splitting them into two separate questions. For example, instead of, "How do you like this cell phone compared with

previous ones that you have owned?" Say, "How do you like this cell phone? Have you owned other cell phones? If so, How did you like it?" This is easier for the interviewee and easier for the interviewer to record.

- Avoid using jargon and language that the interviewee may not understand but would be too embarrassed to admit.
- Avoid leading questions such as, "Why do you like this style of interaction?" If used on its own, this question assumes that the person did like it.
- Be alert to unconscious biases. Be sensitive to your own biases and strive for neutrality in your questions.

Asking colleagues to review the questions and running a pilot study will help to identify problems in advance and gain practice in interviewing.

When planning an interview, think about interviewees who may be reticent to answer questions or who are in a hurry. They are doing ***you*** a favor, so try to make it as pleasant for them as possible and try to make the interviewee feel comfortable. Including the following steps will help you to achieve this (Robson, 1993):

1. An ***Introduction*** in which the interviewer introduces himself and explains why he is doing the interview, reassures **interviewees about the ethical issues**, and asks if they mind being recorded, if appropriate. This should be exactly the same for each interviewee.
2. A ***warmup*** session where easy, non-threatening questions come first. These may include questions about demographic information, such as "Where do you live?"
3. A ***main*** session in which the questions are presented in a logical sequence, with the more difficultones at the end.
4. A ***cool-offperiod*** consisting of a few easy questions (to defuse tension if it has arisen).
5. A ***closing*** session in which the interviewer thanks the interviewee and switches off the recorder or puts her notebook away, signaling that the interview has ended.

The golden rule is to be professional. Here is some further advice about conducting interviews (Robson, 1993):

- Dress in a similar way to the interviewees if possible. If in doubt, dress neatly and avoid standing out.
- Prepare an informed consent form and ask the interviewee to sign it.
- If you are recording the interview, which is advisable, make sure your equipment works in advance and you know how to use it.
- Record answers exactly; do not make cosmetic adjustments, correct, or change answers in any way.

13.2.2 Unstructured interviews

Open-ended or unstructured interviews are at one end of a spectrum of how much control the interviewer has on the process. They are more like conversations that focus on a particular topic and may often go into considerable depth. Questions posed by the interviewer are open, meaning that the format and content of answers is not predetermined. The interviewee is free to answer as fully or as briefly as she wishes. Both interviewer and interviewee can steer the interview. Thus one of the skills necessary for this type of interviewing is to make sure that answers to relevant questions are obtained. It is therefore advisable to be organized and have a plan of the main things to be covered. Going in without an agenda to accomplish a goal is not advisable, and should not to be confused with being open to new information and ideas.

A benefit of unstructured interviews is that they generate rich data. Interviewees often mention things that the interviewer may not have considered and can be further explored. But this benefit often comes at a cost. A lot of unstructured data is generated, which can be very time-consuming and difficult to analyze. It is also impossible to replicate the process, since each interview takes on its own format. Typically in evaluation, there is no attempt to analyze these interviews in detail. Instead, the evaluator makes notes or records the session and then goes back later to note the main issues of interest.

The main points to remember when conducting an unstructured interview are:

- Make sure you have an interview agenda that supports the study goals and questions (identified through the DECIDE framework).
- Be prepared to follow new lines of enquiry that contribute to your agenda.
- Pay attention to ethical issues, particularly the need to get informed consent.
- Work on gaining acceptance and putting the interviewees at ease. For example, dress as they do and take the time to learn about their world.
- Respond with sympathy if appropriate, but be careful not to put ideas into the heads of respondents.
- Always indicate to the interviewee the beginning and end of the interview session.
- Start to order and analyze your data as soon as possible after the interview.

ACTIVITY 13.1 Ananova is a virtual news reporter created by the British Press Association on the website www.ananova.com, which is similar to the picture in Figure 13.1. Viewers who wish to hear Ananova report the news must select from the menu beneath her picture and must have downloaded software that enables them to receive streaming video. Those who wish to read text may do so.

The idea is that Ananova is a life-like, i.e., an 'anthropomorphic' news presenter. She is designed to speak, move her lips, and blink, and she has some human facial expressions. She reads news edited from news reports. Ananova's face, her voice tone, her hair, in fact everything about her was tested with users before the site was launched so that she would appeal to as many users as possible. She is fashionable and looks as though she is in her twenties or

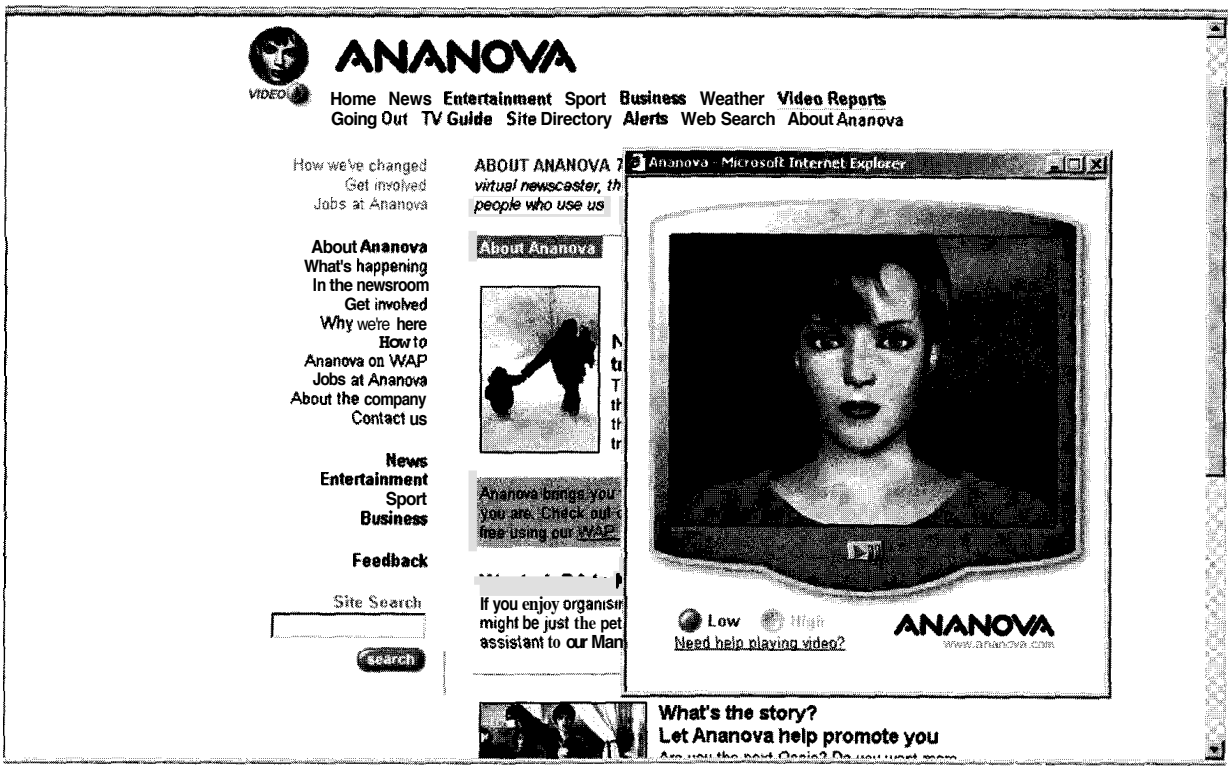


Figure 13.1 Ananova.com showing **Ananova**, a virtual news presenter.

early thirties — presumably the age that market researchers determined fits the profile of the majority of users — and she is also designed to appeal to older people too.

To see **Ananova** in action, go to the website (www.annanova.com) and follow the directions for downloading the software. Alternatively you can do the activity by just looking at the figure and thinking about the questions.

(a) Suggest unstructured interview questions that seek opinions about whether **Ananova** improves the quality of the news service.

(b) Suggest ways of collecting the interview data.

(c) Identify practical and ethical issues that need to be considered.

Comment

(a) Possible questions include: Do you think **Ananova** reading the news is good? Is it better than having to read it yourself from a news bulletin? In what ways does having **Ananova** read the news influence your satisfaction with the service?

(b) Taking notes might be cumbersome and distracting to the interviewee, and it would be easy to miss important points. An alternative is to audio record the session. Video recording is not needed as it isn't necessary to see the interviewee. However, it would be useful to have a camera at hand to take shots of the interface in case the interviewee wanted to refer to aspects of **Ananova**.

(c) The obvious practical issues are obtaining a cassette recorder, finding participants, scheduling times for the interviews and finding a quiet place to conduct them. Having

a computer available for the interviewee to refer to is important. The ethical issues include telling the interviewees why you are doing the interviews and what you will do with the information, and guaranteeing them anonymity. An informed consent form may be needed.

13.2.3 Structured interviews

Structured interviews pose predetermined questions similar to those in a questionnaire (see Section 13.3). Structured interviews are useful when the study's goals are clearly understood and specific questions can be identified. To work best, the questions need to be short and clearly worded.,Responses may involve selecting from a set of options that are read aloud or presented on paper. The questions should be refined by asking another evaluator to review them and by running a small pilot study. Typically the questions are ***closed,*** which means that they require a precise answer. The same questions are used with each participant so the study is standardized.

13.2.4 Semi-structured interviews

Semi-structured interviews combine features of structured and unstructured interviews and use both closed and open questions. For consistency the interviewer has a basic script for guidance, so that the same topics are covered with each interviewee. The interviewer starts with preplanned questions and then probes the interviewee to say more until no new relevant information is forthcoming. For example:

> Which websites do you visit most frequently? ***<Answer>*** Why? ***<Answer mentions several but stresses that she prefers hottestmusic.com>*** And why do you like it? ***<Answer>*** Tell me more about x? ***<silence, followed by an answer>*** Anything else? ***<Answer>*** Thanks. Are there any other reasons that you haven't mentioned?

It is important not to preempt an answer by phrasing a question to suggest that a particular answer is expected. For example, "You seemed to like this use of color . . ." assumes that this is the case and will probably encourage the interviewee to answer that this is true so as not to offend the interviewer. Children are particularly prone to behave in this way. The body language of the interviewer,for example, whether she is smiling,scowling,looking disapproving,etc.,can have a strong influence.

Also the interviewer needs to accommodate ***silences*** and not to move on too quickly. Give the person time to speak. ***Probes*** are a device for getting more information, especially neutral probes such as, "Do you want to tell me anything else?" You may also ***prompt*** the person to help her along. For example, if the interviewee is talking about a computer interface but has forgotten the name of a key menu item, you might want to remind her so that the interview can proceed productively. However, semi-structured interviews are intended to be broadly replicable, so probing and prompting should aim to help the interview along without introducing bias.

ACTIVITY 13.2 Write a semi-structured interview script to evaluate whether receiving news from Ananova is appealing and whether Ananova's presentation is realistic. Show two of your peers the

Ananova.com website or Figure 13.1. Then ask them to comment on your interview script. Refine the questions based on their comments.

Comment

You can use questions that have a predetermined set of answer choices. These work well for fast interviews when the range of answers is known, as in the airport studies where people tend to be in a rush. Alternatively, open-ended questions can also be used if you want to explore the range of opinions.

Some questions that you might ask include:

- Have you seen Ananova before?
- Would you like to receive news from Ananova?
- Why?
- In your opinion, does Ananova look like a real person?

Some of the questions in Exercise 13.2 have a predetermined range of answers, such as "yes," "no," "maybe." Others, such as the one about interviewees' attitudes, do not have an easily predicted range of responses. But it would help us in collecting answers if we list possible responses together with boxes that can just be checked (i.e., ticked). Here's how we could convert the questions from Activity 13.2.

- Have you seen Ananova before? (Explore previous knowledge)
  Interviewer checks box ☐ *Yes* ☐ *No* ☐ *Don't remember/know*
- Would you like to receive news from Ananova? (Explore initial reaction, then explore the response)
  Interviewer checks box ☐ *Yes* ☐ *No* ☐ *Don't know*
- Why?
  If response is "Yes" or "No," interviewer says, "Which of the following statements represents your feelings best?"
  For "Yes," Interviewer checks the box
  - ☐ *I don't like typing*
  - ☐ *This is fun/cool*
  - ☐ *I've never seen a system like this before*
  - ☐ *It's going to be the way of the future*
  - ☐ *Another reason (Interviewer notes the reason)*

  For "No," Interviewer checks the box
  - ☐ *I don't like speech systems*
  - ☐ *I don't like systems that pretend to be people*
  - ☐ *It's faster to read*
  - ☐ *I can't control the pace of presentation*
  - ☐ *I can't be bothered to download the software*
  - ☐ *Another reason (Interviewer notes the reason)*
- In your opinion, does Ananova look like a real person?
  Interviewer checks box
  - ☐ *Yes, she looks like a real person*
  - ☐ *No, she doesn't look like a real person*

As you can probably guess, there are problems deciding on the range of possible answers. Maybe you thought of other ones. In order to get a good range of answers for the second question, a large number of people would have to be interviewed before the questionnaire is constructed to identify all the possible answers and then those could be used to determine what should be offered.

ACTIVITY 13.3 Write three or four semi-structured interview questions to find out if **Ananova** is popular with your friends. Make the questions general.

Comment Here are some suggestions:

(a) Would you listen to the news using **Ananova**?
If yes, then ask, why?
If no, then ask, why not?

(b) Is Ananova's appearance attractive to you?
If yes, then say, Tell me more, what did you like?
If no, then *say, What* don't you find attractive?

(c) Is there anything else you want to say about **Ananova**?

ACTIVITY 13.4 Prepare the **full** interview script to evaluate **Ananova,** including a description of why you are doing the interview, and an informed consent form, and the exact questions. Use the DECIDE framework for guidance. Practice the interview on your own, audiotape yourself, and then listen to it and review your performance. Then interview two peers and be reflective. What did you learn from the experience?

Comment You probably found it harder than you thought to interview smoothly and consistently. Did you notice an improvement when you did the second interview? Were some of the questions poorly worded. Piloting your interview often reveals poor or ambiguous questions that you then have a chance to refine before holding the first proper interview.

13.2.5 Group interviews

One form of group interview is the focus group that is frequently used in marketing, political campaigning, and social sciences research. Normally three to 10 people are involved. Participants are selected to provide a representative sample of typical users; they normally share certain characteristics. For example, in an evaluation of a university **website,** a group of administrators, faculty, and students may be called to form three separate focus groups because they use the web for different purposes.

The benefit of a focus group is that it allows diverse or sensitive issues to be raised that would otherwise be missed. The method assumes that individuals develop opinions within a social context by talking with others. Often questions posed to focus groups seem deceptively simple but the idea is to enable people to put forward their own opinions in a supportive environment. A preset agenda is developed to guide the discussion but there is sufficient flexibility for a facilitator to

follow unanticipated issues as they are raised. The facilitator guides and prompts discussion and skillfully encourages quiet people to participate and stops verbose ones from dominating the discussion. The discussion is usually recorded for later analysis in which participants my be invited to explain their comments more fully.

Focus groups appear to have high validity because the method is readily understood and findings appear believable (Marshall and Rossman, 1999). Focus groups are also attractive because they are low-cost, provide quick results, and can easily be scaled to gather more data. Disadvantages are that the facilitator needs to be skillful so that time is not wasted on irrelevant issues. It can also be difficult to get people together in a suitable location. Getting time with any interviewees can be difficult, but the problem is compounded with focus groups because of the number of people involved. For example, in a study to evaluate a university website the evaluators did not expect that getting participants would be a problem. However, the study was scheduled near the end of a semester when students had to hand in their work, so strong incentives were needed to entice the students to participate in the study. It took an increase in the participation fee and a good lunch to convince students to participate.

13.2.6 Other sources of interview-like feedback

Telephone interviews are a good way of interviewing people with whom you cannot meet. You cannot see body language, but apart from this telephone interviews have much in common with face-to-face interviews.

Online interviews, using either asynchronous communication as in email or synchronous communication as in chats, can also be used. For interviews that involve sensitive issues, answering questions anonymously may be preferable to meeting face to face. If, however, face to face meetings are desirable but impossible because of geographical distance, video-conferencing systems can be used (but remember the drawbacks discussed in Chapter 4). Feedback about a product can also be obtained from customer help lines, consumer groups, and online customer communities that provide help and support.

At various stages of design, it is useful to get quick feedback from a few users. These short interviews are often more like conversations in which users are asked their opinions. Retrospective interviews can be done when doing field studies to check with participants that the interviewer has correctly understood what was happening.

DILEMMA What They Say and What They Do!

What users say isn't always what they do. People sometimes give the answers that they think show them in the best light, or they may just forget what happened or how long they spent on a particular activity.

So, can evaluators believe all the responses they get? Are the respondents giving "the truth" or are they simply giving the answers that they think the evaluator wants to hear?

It isn't possible to avoid this behavior, but it is important to be aware of it and to reduce such biases by getting a large number of participants or by using a combination of techniques. Questions that suggest particular responses should also be avoided.

13.2.7 Data analysis and interpretation

Analysis of unstructured interviews can be time-consuming, though their contents can be rich. Typically each interview question is examined in depth in a similar way to observation data discussed in Chapter 12. A coding form may be developed, which may be predetermined or may be developed during data collection as evaluators are exposed to the range of issues and learn about their relative importance. Alternatively, comments may be clustered along themes and anonymous quotes used to illustrate points of interest. Tools such a NUDIST and Ethnograph can be useful for qualitative analyses as mentioned in Chapter 12. Which type of analysis is done depends on the goals of the study, as does whether the whole interview is transcribed, only part of it, or none of it. Data from structured interviews is usually analyzed quantitatively as in questionnaires which we discuss next.

13.3 Asking users: questionnaires

Questionnaires are a well-established technique for collecting demographic data and users' opinions. They are similar to interviews and can have ***closed*** or open questions. Effort and skill are needed to ensure that questions are clearly worded and the data collected can be analyzed efficiently. Questionnaires can be used on their own or in conjunction with other methods to clarify or deepen understanding. In the HutchWorld study discussed in Chapter 10, for example, you read how questionnaires were used along with observation and usability testing. The methods and questions used depends on the context, interviewees and so on.

The questions asked in a questionnaire, and those used in a structured interview are similar, so how do you know when to use which technique? One advantage of questionnaires is that they can be distributed to a large number of people. Used in this way, they provide evidence of wide general opinion. On the other hand, structured interviews are easy and quick to conduct in situations in which people will not stop to complete a questionnaire.

13.3.1 Designing questionnaires

Many questionnaires start by asking for basic demographic information (e.g., gender, age) and details of user experience (e.g., the time or number of years spent using computers, level of expertise, etc.). This background information is useful in finding out the range within the sample group. For instance, a group of people who are using the web for the first time are likely to express different opinions to another group with five years of web experience. From knowing the sample range, a designer might develop two different versions or veer towards the needs of one of the groups more because it represents the target audience.

Following the general questions, specific questions that contribute to the evaluation goal are asked. If the questionnaire is long, the questions may be subdivided into related topics to make it easier and more logical to complete.

Box 13.1 contains an excerpt from a paper questionnaire designed to evaluate users' satisfaction with some specific features of a prototype website for career changers aged 34–59 years.

BOX 13.1 An Excerpt from A User Satisfaction Questionnaire Used to Evaluate A Website for Career Changers *(Andrews et al., 2001)*

Notice that in the following excerpt users are asked to circle appropriate responses, and check the box that most closely describes their opinion: these are commonly used techniques. Fewer than fifty participants were involved in this study, so inviting them to write on open-ended comment suggesting recommendations for change was manageable. It would have been difficult to collect this information with closed questions, since good suggestions would undoubtedly have been missed because the evaluator would probably not have thought to ask about them.

Participant #: _____

Please circle the most appropriate selection:

Age Range: 34–39 40–49 50–59

Gender: Male Female

Career-Change Status: Exploring In-Progress Completed

Internet/Web Experience

Research, Information Gathering Daily Weekly Monthly Never

Bulletin Board Posting Daily Weekly Monthly Never

Chat Room Usage Daily Weekly Monthly Never

Please rate (i.e., check the box to show) agreement or disagreement with the following statements:

Question	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The navigation language on the links is clear and easy to understand					
The website site contains information that would be useful to me					
Information on the website is easy to find					
The "Center Design" presents information in an aesthetically pleasing manner					
The website pages are confusing and difficult to read					
I prefer darker colors to lighter colors for display					
It is apparent from the first website page (homepage) what the purpose of the website is.					

Please add any recommendations for changes to the overall design, language or navigation of the website on the back of this paper.

Thanks for your participation in the testing of this prototype.

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

- Make questions clear and specific.
- When possible, ask closed questions and offer a range of answers.
- Consider including a "no-opinion" option for questions that seek opinions.
- Think about the ordering of questions. The impact of a question can be influenced by question order. General questions should precede specific ones.
- Avoid complex multiple questions.
- When scales are used, make sure the range is appropriate and does not overlap.
- Make sure that the ordering of scales (discussed below) is intuitive and consistent, and be careful with using negatives. For example, it is more intuitive in a scale of 1 to 5 for 1 to indicate low agreement and 5 to indicate high agreement. Also be consistent. For example, avoid using 1 as low on some scales and then as high on others. A subtler problem occurs when most questions are phrased as positive statements and a few are phrased as negatives. However, advice on this issue is more controversial as some evaluators argue that changing the direction of questions helps to check the users' intentions. Scales such as those used in Box 13.1 are also preferred by some evaluators.
- Avoid jargon and consider whether you need different versions of the questionnaire for different populations.
- Provide clear instructions on how to complete the questionnaire. For example, if you want a check put in one of the boxes, then say so. Questionnaires can make their message clear with careful wording and good typography.
- A balance must be struck between using white space and the need to keep the questionnaire as compact as possible. Long questionnaires cost more and deter participation.

13.3.2 Question and response format

Different types of questions require different types of responses. Sometimes discrete responses are required, such as "Yes" or "No." For other questions it is better to ask users to locate themselves within a range. Still others require a single preferred opinion. selecting the most appropriate makes it easier for respondents to be able to answer. Furthermore, questions that accept a specific answer can be categorized more easily. Some commonly used formats are described below.

Check boxes and ranges

The range of answers to demographic questionnaires is predictable. Gender, for example, has two options, male or female, so providing two boxes and asking respondents to check the appropriate one, or circle a response, makes sense for collecting this information (as in Box 13.1). 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 (Box 13.1). A common design error arises when the ranges overlap. For example, specifying two ranges as 15–20, 20–25 will cause confusion: which box do people who are 20 years old check? Making the ranges 14–19, 20–24 avoids this problem.

A frequently asked question about ranges is whether the interval must be equal in all cases. The answer is that it depends on what you want to know. For example, if you want to collect information for the design of an e-commerce site to sell life insurance, the target population is going to be mostly people with jobs in the age range of, say, 21-65 years. You could, therefore, have just three ranges: under 21, 21–65 and over 65. In contrast, if you are interested in looking at ten-year cohort groups for people over 21 the following ranges would be best: under 21, 22–31, 32–41, etc.

There are a number of different types of rating scales that can be used, each with its own purpose (see Oppenheim, 1992). Here we describe two commonly used scales, Likert and semantic differential scales.

The purpose of these 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 about things, e.g. how easy, how usable etc., and therefore are important for usability studies.

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 the most commonly used scales because identifying suitable statements that respondents will understand 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 as in the HutchWorld evaluation described in Chapter 10. For example, users' opinions about the use of color in a website could be evaluated with a Likert scale using a range of numbers (1) or with words (2):

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

1	2	3	4	5
☐	☐	☐	☐	17

(2) The use of color is excellent:

strongly agree	agree	OK	disagree	strongly disagree
☐	☐	☐	☐	☐

Below are some steps for designing Likert scales:

- Gather a pool of short statements about the features of the product that are to be evaluated e.g., "This control panel is easy to use." A brainstorming

session with peers in which you examine the product to be evaluated is a good way of doing this.

- Divide the items into groups with about the same number of positive and negative statements in each group. Some evaluators prefer to have all negative or all positive questions, while others use a mix of positive and negative questions, as we have suggested here. Deciding whether to phrase the questionnaire positively or negatively depends partly on the complexity of the questionnaire and partly on the evaluator's preferences. The designers of **QUIS** (Box 13.2) (Chin et al., 1988), for example, decided not to mix negative and positive statements because the questionnaire was already complex enough without forcing participants to pay attention to the direction of the argument.
- Decide on the scale. **QUIS** (Box 13.2) uses a 9-point scale, and because it is a general questionnaire that will be used with a wide variety of products it also includes N/A (not applicable,) as a category. Many questionnaires use 7- or 5-point scales and there are also 3-point scales. Arguments for the number of points go both ways. Advocates of long scales argue that they help to show discrimination, as advocated by the **QUIS** team (Chin et al., 1988). 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 counter-argument 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 the scale should have an even or odd number of points. An odd number provides a clear central point. On the

BOX 13.2 QUIS, Questionnaire for User Interaction Satisfaction

The Questionnaire for User Interaction Satisfaction (QUIS) developed by the University of Maryland Human-Computer Interaction Laboratory is one of the most widely used questionnaires for evaluating interfaces (Chin et al., 1988; Shneiderman, 1998a). Although developed for evaluating user satisfaction, it is frequently applied to other aspects of interaction design. An advantage of this questionnaire is that it has gone through many cycles of refinement and has been used for hundreds of evaluation studies, so it is well tried and tested. The questionnaire consists of the following 12 parts that can be used in total or in parts:

- system experience (i.e., time spent on this system)
- past experience (i.e., experience with other systems)
- overall user reactions
- screen design
- terminology and system information
- learning (i.e., to operate the system)
- system capabilities (i.e., the time it takes to perform operations)
- technical manuals and online help
- online tutorials
- multimedia
- teleconferencing
- software installation

Notice that the third part of QUIS assesses users' overall reactions. Evaluators often use this part on its own because it is short so people are likely to respond.

other hand, an even number forces participants to make a decision and prevents them from sitting on the fence.

- Select items for the final questionnaire and reword as necessary to make them clear.

Semantic differential scales

Semantic differential scales are used less frequently than Likert scales. They explore a range of bipolar attitudes about a particular item. Each pair of attitudes is represented as a pair of adjectives. The participant is asked to place a cross in one of a number of positions between the two extremes to indicate agreement with the poles, as shown in Figure 13.2. The score for the evaluation is found by summing the scores for each bipolar pair. Scores can then be 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.

Instructions: for each pair of adjectives, place a cross at the point between them that reflects the extent to which you believe the adjectives describe the home page. You should place ***only one cross*** between the marks on each line.

Attractive		Ugly
Clear		Confusing
Dull		Colorful
Exciting		Boring
Annoying		Pleasing
Helpful		Unhelpful
Poor		Well designed

Figure 13.2 An example of a semantic differential scale.

ACTIVITY 13.5 Spot the four poorly designed features in Figure **13.3.**

Comment

Some of the features that could be improved include:

- Request for exact age. Many people prefer not to give this information and would rather position themselves in a range.
- Years of experience is indicated with overlapping scales, i.e., <1, 1–3, 3–5, etc. How do you answer if you have 1, 3, or 5 years of experience?
- The questionnaire doesn't tell you whether you should check one, two, or as many boxes as you wish.
- The space left for people to write their own information is too small, and this will annoy them and deter them from giving their opinions.

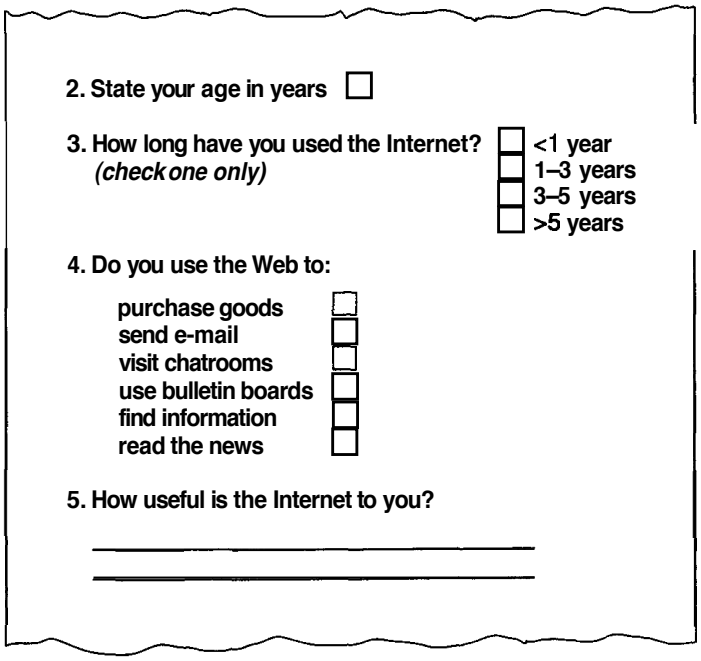


Figure 13.3 A questionnaire with poorly designed features.

13.3.3 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 tend to use small numbers of participants, often fewer than twenty users. One hundred percent completion rates often are achieved with these small samples, but with larger, more remote populations, ensuring that surveys are returned is a well-known problem. Forty percent return is generally acceptable for many surveys but much lower rates are common.

Some ways of encouraging a good response include:

- Ensuring the questionnaire is well designed so that participants do not get annoyed and give up.
- Providing a short overview section, as in QUIS (Box 13.2), and telling respondents to complete just the short version if they do not have time to complete the whole thing. This ensures that you get something useful returned.
- Including a stamped, self-addressed envelope for its return.
- Explaining why you need the questionnaire to be completed and assuring anonymity.
- Contacting respondents through a follow-up letter, phone call or email.
- Offering incentives such as payments.

13.3.4 Online questionnaires

Online questionnaires are becoming increasingly common because they are effective for reaching large numbers of people quickly and easily. There are two types: email and web-based. The main advantage of email is that you can target specific users. However, email questionnaires are usually limited to text, whereas web-based questionnaires are more flexible and can include check boxes, pull-down and pop-up menus, help screens, and graphics (Figure 13.4). web-based questionnaires can also provide immediate data validation and can enforce rules such as select only one response, or certain types of answers such as numerical, which cannot be done in email or with paper. Other advantages of online questionnaires include (Lazar and Preece, 1999):

- Responses are usually received quickly.
- Copying and postage costs are lower than for paper surveys or often non-existent.
- Data can be transferred immediately into a database for analysis.
- The time required for data analysis is reduced.
- Errors in questionnaire design can be corrected easily (though it is better to avoid them in the first place).

A big problem with web-based questionnaires is obtaining a random sample of respondents. Few other disadvantages have been reported with online questionnaires, but there is some evidence suggesting that response rates may be lower online than with paper questionnaires (Witmer et al., 1999).

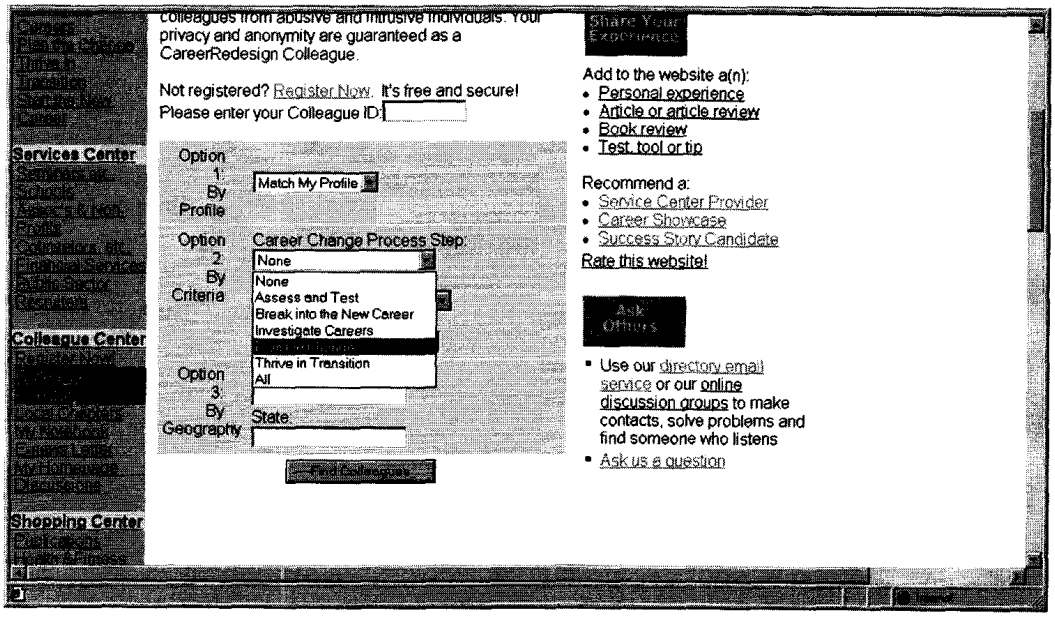


Figure 13.4 An excerpt from a web-based questionnaire showing pull-down menus.

Developing a web-based questionnaire

Developing a successful web-based questionnaire involves designing it on paper, developing strategies for reaching the target population, and then turning the paper version into a web-based version (Lazar and Preece, 1999).

It is important to devise the questionnaire on paper first, following the general guidelines introduced above, such as paying attention to the clarity and consistency of the questions, questionnaire layout, and so on. Only once the questionnaire has been reviewed and the questions refined adequately should it be translated into a web-based version. If reaching your target population is an issue, e.g., if some of them may not have access to the web, the paper version may be administered to them, but be careful to maintain consistency between the web-based version and the original paper version.

Identifying a random sample of a population so that the results are indicative of the whole population may be difficult, if not impossible, to achieve especially if the size and demography of the population is not known, as is often the case in Internet research. This has been a criticism of several online surveys including Georgia Tech's GVU survey, one of the first online surveys. This survey collects demographic and activity information from Internet users and has been distributed twice yearly since 1994. The policy that GVU employs to deal with this difficult sampling issue is to make as many people aware of the GVU survey as possible so that a wide variety of participants are encouraged to participate. However, even these efforts do not avoid biased sampling, since participants are self-selecting. Indeed, some survey experts are vehemently opposed to such methods and instead propose using national census records to sample offline (Nie & Ebring, 2000). In some countries, web-based questionnaires are used in conjunction with television to elicit viewers' opinions of programs and political events, and many such questionnaires now say that their results are "not scientific" when they cite them, meaning that unbiased sampling was not done. A term that is gaining popularity is convenience sampling, which is another way of saying that the sample includes those who were available rather than those selected using scientific sampling.

Turning the paper questionnaire into a web-based version requires four steps.

1. Produce an error-free interactive electronic version from the original paper-based one. This version should provide clear instructions and be free of input errors. For example, if just one box should be checked, the other attempts should be rejected automatically. It may also be useful to embed feedback and pop-up help within the questionnaire.
2. Make the questionnaire accessible from all common browsers and readable from different-size monitors and different network locations. Specialized software or hardware should be avoided. The need to download software also deters novice users and should be avoided.
3. Make sure information identifying each respondent will be captured and stored confidentially because the same person may submit several completed surveys. This can be done by recording the Internet domain name or the IP address of the respondent, which can then be transferred directly to a

database. However, this action could infringe people's privacy and the legal situation should be checked. Another way is to access the transfer and referrer logs from the web server, which provide information about the domains from which the web-based questionnaire was accessed. Unfortunately, people can still send from different accounts with different IP addresses, so additional identifying information may also be needed.

4. User-test the survey with pilot studies before distributing.

Commercial questionnaires are becoming available via the Internet. Two examples are SUMI and MUMMS, which are briefly discussed in Box **13.3.**

BOX 13.3 Questionnaire Tools

SUMI (Software Usability Measurement Inventory) was developed in the early 1990s as part of a European project. The aim was to develop a standardized tool to evaluate users' reactions to a piece of software. More recently a new version has been developed, known as MUMMS (Measuring the Usability of Multi-Media Systems), that, as the name implies, is geared more towards current software in which multimedia is assumed to be a component. This questionnaire focuses on five concepts:

- how much the product captures the user's emotional responses
- how much the user feels in control of the software
- the degree to which the users can achieve their goals using the software
- the extent to which the product seems to assist the user
- the ease with which the user can learn to use the product

The developers are also planning to include a new concept that they are calling "excitement." This would address two kinds of user experience goals (emotional responses and excitement) and four usability goals. More information about SUMI and MUMMS can be found at www.ucc.ie/hfrg/questionnaires/

13.3.5 Analyzing questionnaire data

Having collected a set of questionnaire responses, you need to know what to do with the data. The first step is to identify any trends or patterns. Using a spreadsheet like Excel to hold the data can help in this initial analysis. Often only simple statistics are needed such as the number or percentage of responses in a particular category. If the number of participants is small, under ten for example, giving actual numbers is more honest, but for larger numbers of responses percentages are useful for standardizing the data, particularly if you want to compare two or more sets of responses. Bar charts can also be used to display data graphically. More advanced statistical techniques such as cluster analysis can also be used to show whether there is a relationship between question responses.

13.4 Asking experts: inspections

Sometimes users are not easily accessible or involving them is too expensive or takes too long. In such circumstances, experts or combinations of experts and users can

provide feedback. Various inspection techniques began to be developed as alternatives to usability testing in the early 1990s. These included various kinds of expert evaluations or ***reviews,*** such as heuristic evaluations and walkthroughs, in which experts inspect the human-computer interface and predict problems users would have when interacting with it. Typically these techniques are relatively inexpensive and easy to learn as well as being effective, which makes them appealing. They are similar to some software engineering practices where code and other types of inspections have been conducted for years. In addition, they can be used at any stage of a design project, including early design before well-developed prototypes are available.

13.4.1 Heuristic evaluation

Heuristic evaluation is an informal usability inspection technique developed by Jakob Nielsen and his colleagues (Nielsen, 1994a) in which experts, guided by a set of usability principles known as ***heuristics,*** evaluate whether user-interface elements, such as dialog boxes, menus, navigation structure, online help, etc., conform to the principles. These heuristics closely resemble the high-level design principles and guidelines discussed in Chapters 1 and 8, e.g., making designs consistent, reducing memory load, and using terms that users understand. When used in evaluation, they are called heuristics. The original set of heuristics was derived empirically from an analysis of 249 usability problems (Nielsen, 1994b). We list the latest here (also in Chapter 1), this time expanding them to include some of the questions addressed when doing evaluation:

- ***Visibility of system status***
  Are users kept informed about what is going on?
  Is appropriate feedback provided within reasonable time about a user's action?
- ***Match between system and the real world***
  Is the language used at the interface simple?
  Are the words, phrases and concepts used familiar to the user?
- ***User control and freedom***
  Are there ways of allowing users to easily escape from places they unexpectedly find themselves in?
- ***Consistency and standards***
  Are the ways of performing similar actions consistent?
- ***Help users recognize, diagnose, and recover from errors***
  Are error messages helpful?
  Do they use plain language to describe the nature of the problem and suggest a way of solving it?
- ***Error prevention***
  Is it easy to make errors?
  If so where and why?
- ***Recognition rather than recall***
  Are objects, actions and options always visible?

- *Flexibility and efficiency of use*
  Have accelerators (i.e., shortcuts) been provided that allow more experienced users to carry out tasks more quickly?
- *Aesthetic and minimalist design*
  Is any unnecessary and irrelevant information provided?
- *Help and documentation*
  Is help information provided that can be easily searched and easily followed?

However, some of these core heuristics are too general for evaluating new products coming onto the market and there is a strong need for heuristics that are more closely tailored to specific products. For example, Nielsen (1999) suggests that the following heuristics are more useful for evaluating commercial websites, and makes them memorable by introducing the acronym HOMERUN:

- High-quality content
- Often updated
- Minimal download time
- Ease of use
- Relevant to users' needs
- **Unique to the online medium**
- Netcentric corporate culture

Different sets of heuristics for evaluating toys, WAP devices, online communities, wearable computers, and other devices are needed, so evaluators must develop their own by tailoring Nielsen's heuristics and by referring to design guidelines, market research, and requirements documents. Exactly which heuristics are the best and how many are needed are debatable and depend on the product.

Using a set of heuristics, expert evaluators work with the product role-playing typical users and noting the problems they encounter. Although other numbers of experts can be used, empirical evidence suggests that five evaluators usually identify around 75% of the total usability problems, as shown in Figure 13.5 (Nielsen,

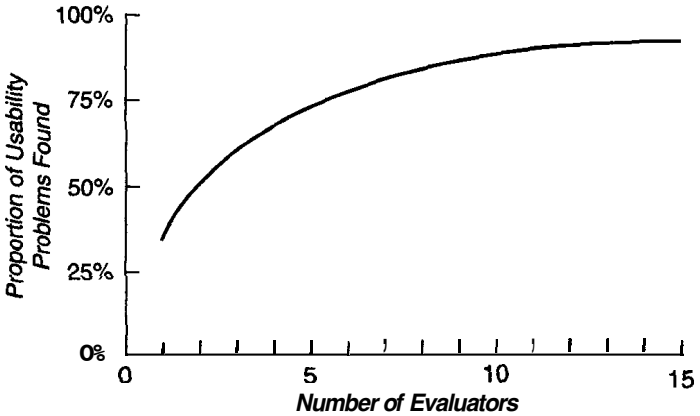
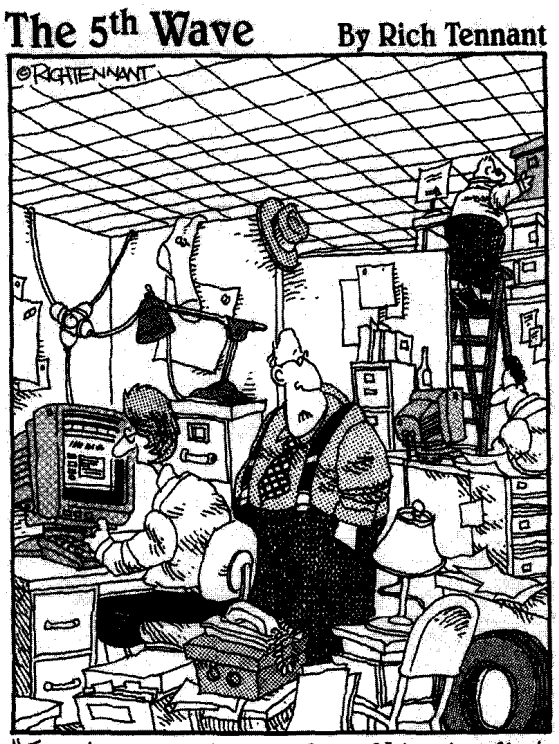


Figure 13.5 Curve showing the proportion of usability problems in an interface found by heuristic evaluation using various numbers of evaluators. The curve represents the average of six case studies of heuristic evaluation.



1994a). However, skillful experts can capture many of the usability problems by themselves, and many consultants now use this technique as the basis for critiquing interactive devices—a process that has become know as an ***expert crit*** in some countries. Because users and special facilities are not needed for heuristic evaluation and it is comparatively inexpensive and quick, it is also known as ***discount evaluation.***

13.4.2 Doing heuristic evaluation

Heuristic evaluation is one of the most straightforward evaluation methods. The evaluation has three stages:

1. ***The briefing session*** in which the experts are told what to do. **A** prepared script is useful as a guide and to ensure each person receives the same briefing.
2. The ***evaluation period*** in which each expert typically spends 1–2 hours ***independently*** inspecting the product, using the heuristics for guidance. The experts need to take ***at least two*** passes through the interface. The ***first pass*** gives a feel for the flow of the interaction and the product's scope. The ***second pass*** allows the evaluator to focus on specific interface ele-

ments in the context of the whole product, and to identify potential usability problems.

If the evaluation is for a functioning product, the evaluators need to have some specific user tasks in mind so that exploration is focused. Suggesting tasks may be helpful but many experts do this automatically. However, this approach is less easy if the evaluation is done early in design when there are only screen mockups or a specification; the approach needs to be adapted to the evaluation circumstances. While working through the interface, specification or mockups, a second person may record the problems identified, or the evaluator may think aloud. Alternatively, she may take notes herself. Experts should be encouraged to be as specific as possible and to record each problem clearly.

3. The *debriefing session* in which the experts come together to discuss their findings and to prioritize the problems they found and suggest solutions.

The heuristics focus the experts' attention on particular issues, so selecting appropriate heuristics is therefore critically important. Even so, there is sometimes less agreement among experts than is desirable, as discussed in the dilemma below.

There are fewer practical and ethical issues in heuristic evaluation than for other techniques because users are not involved. A week is often cited as the time needed to train experts to be evaluators (Nielsen and Mack, 1994), but this of course depends on the person's expertise. The best experts will have expertise in both interaction design and the product domain. Typical users can be taught to do

DILEMMA Problems or False Alarms?

You might think that heuristic evaluation is a panacea for designers, and that it can reveal all that is wrong with a design. However, it has problems. Several independent studies compare heuristic evaluation with other techniques, particularly user testing, indicating that the different approaches often identify *different* problems and that sometimes heuristic evaluation misses severe problems (Karat, 1994). This argues for using complementary techniques. Furthermore, heuristic evaluation should not be thought of as a replacement for user testing.

Another problem that Bill Bailey (2001) warns about is of experts reporting problems that don't exist. In other words, some of the experts' predictions are wrong. Bailey cites analyses from three published sources showing that about 33% of the problems reported were real usability problems, some of which were serious, others trivial. However, the heuristic evaluators missed about 21% of users' problems. Furthermore, about 43% of the problems identified by the experts were *not* problems at all; they were false alarms! Bailey points out that if we do the arithmetic and round up the numbers, what this comes down to is that only about *half* the problems identified are true problems. "More specifically, for every true usability problem identified, there will be a little over one false alarm (1.2) and about one half of one missed problem (0.6). If this analysis is true, heuristic evaluators tend to identify more false alarms and miss more problems than they have true hits."

How can the number of false alarms or missed serious problems be reduced? Checking that experts really have the expertise that they claim would help but how can you do this? One way to over come biases is to have several evaluators. This helps to reduce the impact of one person's bias of poor performance. Using heuristic evaluation along with user testing and other techniques is also a good idea.

heuristic evaluation, although there have been claims that it is not very successful (Nielsen, 1994a). However, some closely related methods take a team approach that involves users (Bias, 1994).

13.4.3 Heuristic evaluation of websites

In this section we examine heuristics for evaluating websites. We begin by discussing MEDLINEplus, a medical information website created by the National Library of Medicine (NLM) to provide health information for patients, doctors, and researchers (Cogdill, 1999). The home page and two other screens are shown in Figures 13.6–13.8.

In 1999 usability consultant Keith Cogdill was commissioned by NLM to evaluate MEDLINEplus. Using a combination of his own knowledge of the users' tasks, problems that had already been reported by users, and advice from documented sources (Shneiderman, 1998a; Nielsen, 1993; Dumas and Redish, 1999), Cogdill identified the seven heuristics listed below. Some of the heuristics resemble Nielsen's original set, but have been tailored for evaluating MEDLINEplus.

- ***Internal consistency.***
  The user should not have to speculate about whether different phrases or actions carry the same meaning.

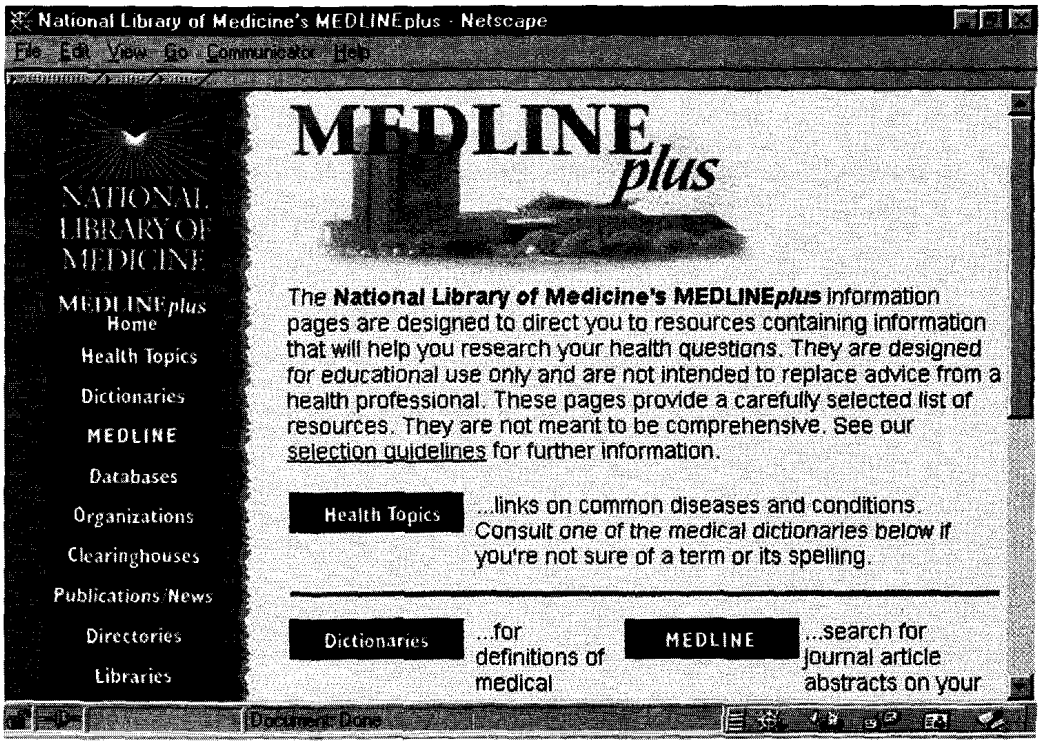


Figure 13.6 Home **page** of **MEDLINEplus.**

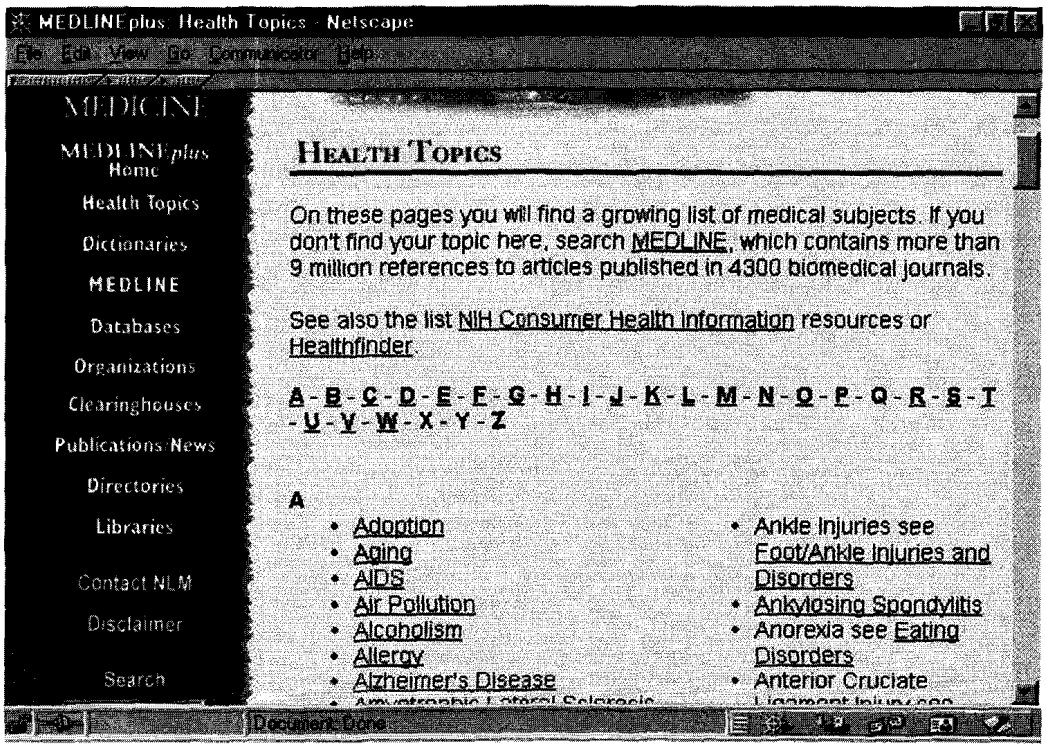


Figure 13.7 Clicking Health Topics on the home page produced this page.

- ***Simple dialog.***
  The dialog with the user should not include information that is irrelevant, unnecessary, or rarely needed. The dialog should be presented in terms familiar to the user and not be system-oriented.
- ***Shortcuts.***
  The interface should accommodate both novice and experienced users.
- ***Minimizing the user's memory load.***
  The interface should not require the user to remember information from one part of the dialog to another.
- ***Preventing errors.***
  The interface should prevent errors from occurring.
- ***Feedback.***
  The system should keep the user informed about what is taking place.
- ***Internal locus of control.***
  Users who choose system functions by mistake should have an "emergency exit" that lets them leave the unwanted state without having to engage in an extended dialog with the system,

These heuristics were given to three expert evaluators who independently evaluated MEDLINEplus. Their comments were then compiled and a meeting was

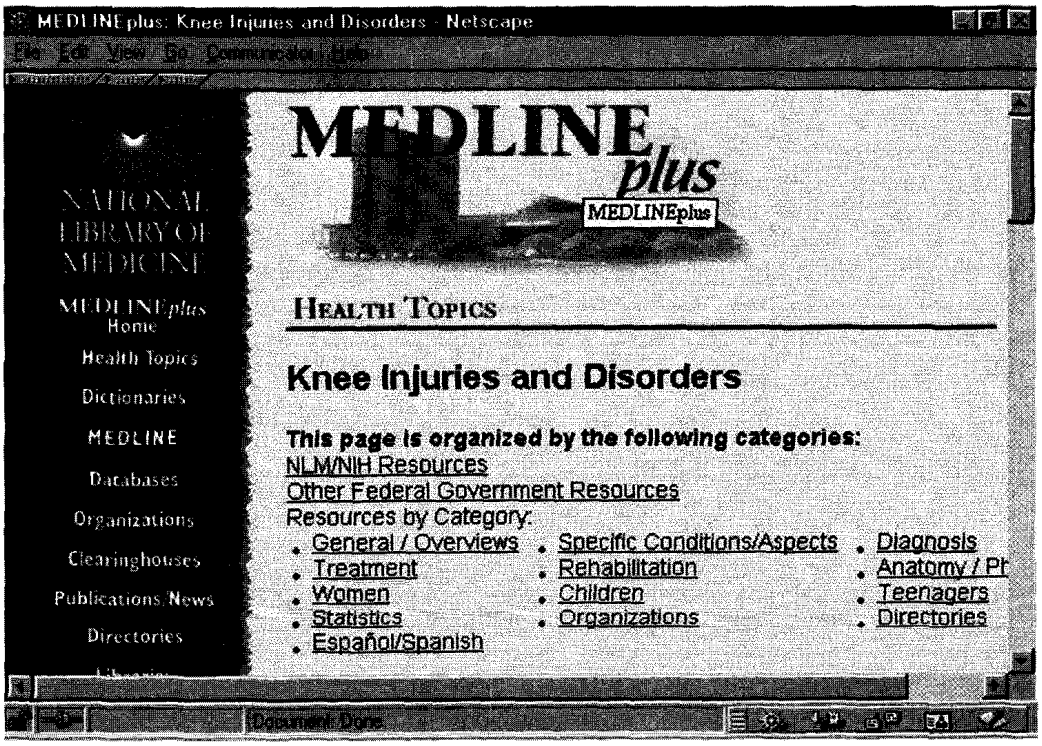


Figure 13.8 Categories of links within Health Topics for knee injuries.

called to discuss their findings and suggest strategies for addressing problems. The following points were among their findings:

- ***Layout.***
  All pages within MEDLINEplus have a relatively uncomplicated vertical design. The home page is particularly compact, and all pages are well suited for printing. The use of graphics is conservative, minimizing the time needed to download pages.
- ***Internal consistency.***
  The formatting of pages and presentation of the logo are consistent across the website. Justification of text, fonts, font sizes, font colors, use of terms, and links labels are also consistent.

The experts also suggested improvements, including:

- ***Arrangement of health topics.***
  Topics should be arranged alphabetically as well as in categories. For example, health topics related to cardiovascular conditions could appear together.
- ***Depth of navigation menu.***
  Having a higher "fan-out" in the navigation menu in the left margin would enhance usability. By this they mean that more topics should be listed on the

surface, giving many short menus rather than a few deep ones (see the experiment on breadth versus depth in Chapter 14 which provides evidence to justify this.)

Turning design guidelines into heuristics for the web

The following list of guidelines for evaluating websites was compiled from several sources and grouped into three categories: *navigation, access,* and *information design* (Preece, 2000). These guidelines provide a basis for developing heuristics by converting them into questions.

Navigation One of the biggest problems for users of large websites is navigating around the site. The phrase "lost in cyberspace" is understood by every web user. The following six guidelines (from Nielsen (1998) and others) are intended to encourage good navigation design:

- *Avoid orphan pages i.e. pages that are not connected to the home page, because they lead users into dead ends.*
  Are there any orphan pages? Where do they go to?
- *Avoid long pages with excessive white space that force scrolling.*
  Are there any long pages? Do they have lots of white space or are they full of texts or lists?
- *Provide navigation support, such as a strong site map that is always present* (Shneiderman, 1998b).
  Is there any guidance, e.g. maps, navigation bar, menus, to help users find their way around the site?
- *Avoid narrow, deep, hierarchical menus that force users to burrow deep into the menu structure.*
  Empirical evidence indicates that broad shallow menus have better usability than a few deep menus (Larson and Czerwinski, 1998; Shneiderman, 1998b).
- *Avoid non-standard link colors.*
  What color is used for links? Is it blue or another color? If it is another color, then is it obvious to the user that it is a hyperlink?
- *Provide consistent look and feel for navigation and information design.*
  Are menus used, named, and positioned consistently? Are links used consistently?

Access Accessing many websites can be a problem for people with slow Internet connections and limited processing power. In addition, browsers are often not sensitive to errors in URLs. Nielsen (1998) suggests the following guidelines:

- *Avoid complex URLs.*
  Are the URLs complex? Is it easy to make typing mistakes when entering them?

- *Avoid long download times that annoy users.*
  Are there pages with lots of graphics? How long does it take to download each page?

Information design **Information design (i.e., content comprehension and aesthetics) contributes to users' understanding and impressions of the site as you can see in Activity 13.6.**

ACTIVITY 13.6 Consider the following design guidelines for information design and for each one suggest a question that could be used in heuristic evaluation:

- ***Outdated or incomplete information is to be avoided*** (Nielsen, 1998). It creates a poor impression with users.
- ***Good graphical design is important.*** Reading long sentences, paragraphs, and documents is difficult on screen, so break material into discrete, meaningful chunks to give the website structure (Lynch and Horton, 1999).
- ***Avoid excessive use of color.*** Color is useful for indicating different kinds of information, i.e., cueing (Preece et al., 1994).
- ***Avoid gratuitous use of graphics and animation.*** In addition to increasing download time, graphics and animation soon become boring and annoying (Lynch and Horton, 1999).
- Be ***consistent.*** Consistency both within pages (e.g., use of fonts, numbering, terminology, etc.) and within the site (e.g., navigation, menu names, etc.) is important for usability and for aesthetically pleasing designs.

Comment We suggest the following questions; you may have identified others:

- ***Outdated or incomplete information.***
  Do the pages have dates on them? How many pages are old and provide outdated information?
- ***Good graphical design is important.***
  Is the page layout structured meaningfully? Is there too much text on each page?
- ***Avoid excessive use of color.***
  How is color used? Is it used as a form of coding? Is it used to make the site bright and cheerful? Is it excessive and garish?
- ***Avoid gratuitous use of graphics and animation.***
  Are there any flashing banners? Are there complex introduction sequences? Can they be short-circuited? Do the graphics add to the site?
- ***Be Consistent.***
  Are the same buttons, fonts, numbers, menu styles, etc. used across the site? Are they used in the same way?

ACTIVITY 13.7 Look at the heuristics above and consider how you would use them to evaluate a website for purchasing clothes (e.g., REI.com, which has a home page similar to that in Figure 13.9).

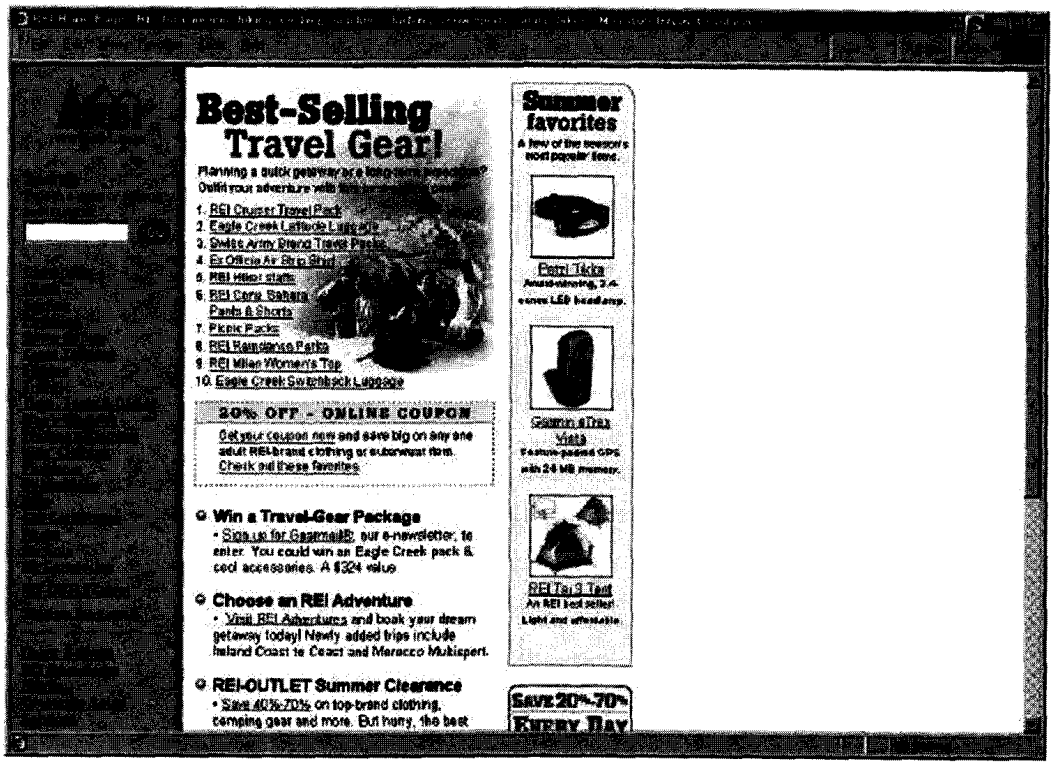


Figure 13.9 The home page is similar to that of REI.com.

While you are doing this activity think about whether the grouping into three categories is useful.

(a) Does it help you focus on what is being evaluated?

(b) Might fewer heuristics be better? Which might be combined and what are the trade-offs?

Comment

(a) Informal evaluation in which the heuristics were categorized suggests that the three categories help evaluators to focus. However, 13 heuristics is still a lot.

(b) Some heuristics can be combined and given a more general description. For example, ***providing navigation support*** and ***avoiding narrow, deep, hierarchical menus*** could be replaced with "help users develop a good mental model," but this is a more abstract statement and some evaluators might not know what is packed into it. Producing questions suitable for heuristic evaluation often results in more of them, so there is a trade-off. An argument for keeping the detail is that it reminds evaluators of the issues to consider. At present, since the web is relatively new, we can argue that such reminders are needed. Perhaps in five years they will not be.

Heuristics for online communities

As we have already mentioned, different combinations and types of heuristics are needed to evaluate different types of applications and interactive products. Another

kind of web application to which heuristics must be tailored is online communities. Here, a key concern is how to evaluate not merely usability but also how well social interaction (i.e., sociability) is supported. This topic has received less attention than the web but the following nine sets of example questions can be used as a starting point for developing heuristics to evaluate online communities (Preece, 2000):

- Sociability: Why should I join this community? (What are the benefits for me? Does the description of the group, its name, its location in the website, the graphics, etc., tell me about the purpose of the group?)
- Usability: How do I join (or leave) the community? (What do I do? Do I have to register or can I just post, and is this a good thing?)
- Sociability: What are the rules? (Is there anything I shouldn't do? Are the expectations for communal behavior made clear? Is there someone who checks that people are behaving reasonably?)
- Usability: How do I get, read and send messages? (Is there support for newcomers? IS it clear what I should do? Are templates provided? Can I send private messages?)
- Usability: Can I do what I want to do easily? (Can I navigate the site? Do I feel comfortable interacting with the software? Can I find the information and people I want?)
- Sociability: Is the community safe? (Are my comments treated with respect? Is my personal information secure? Do people make aggressive or unacceptable remarks to each other?)
- Sociability: Can I express myself as I wish? (Is there a way of expressing emotions, such as using emoticons? Can I show people what I look like or reveal aspects of my character? Can I see others? Can I determine who else is present—perhaps people are looking on but not sending messages?)
- Sociability: Do people reciprocate? (If I contribute will others contribute comments, support and answer my questions?)

  Sociability: Why should I come back? (What makes the experience worthwhile? What's in it for me? Do I feel part of a thriving community? Are there interesting people with whom to communicate? Are there interesting events?)

ACTIVITY 13.8 Go to the communities in REI.com or to another site that has bulletin boards to which customers can send comments. Social interaction **was discussed** in Chapter **4,** and **this exercise** involves picking up some of the concepts discussed there and developing heuristics to evaluate online communities. Before starting you will find it useful to familiarize yourself by carrying out the following:

- read some of the messages
- send a message
- reply to a message
- search for information
- notice how many messages have been sent and how recently

- notice whether you can see the physical relationship between messages easily
- notice whether you can post to people privately using email
- notice whether you can gain a sense of what the other people are like and the emotional content of their messages
- notice whether there is a sense of community and of individuals being present, etc.

Then use the nine questions above as heuristics to evaluate the site:

(a) How well do the questions work as heuristics for evaluating the online community for both usability and sociability issues?

(b) Could these questions form the basis for heuristics for other online communities such as HutchWorld discussed in Chapter 10?

Comment

(a) You probably found that these questions helped focus your attention on the main issues of concern. You may also have noticed that some communities are more like ghost towns than communities; they get very few visitors. Unlike the website evaluation it is therefore important to pay attention to social interaction. A community without people is not a community no matter how good the software is that supports it.

(b) HutchWorld is designed to support social interaction and offers many additional features such as support for social presence by allowing participants to represent themselves as avatars, show pictures of themselves, tell stories, etc. The nine questions above are useful but may need adapting.

13.4.4 Heuristics for other devices

The examples in the previous activities start to show how heuristics can be tailored for specific applications. However, some products are even more different than those from the desktop world of the early 1990s that gave rise to Nielsen's original heuristics. For example, computerized toys are being developed that motivate, entice and challenge, in innovative ways. Handheld devices sell partly on size, color and other aesthetic qualities — features that can have a big impact on the user experience but are not covered by traditional heuristics. Little research has been done on developing heuristics for these products, but Activity 13.9 will start you thinking about them.

ACTIVITY 13.9 Allison Druin works with children to develop web applications and computerized toys (Druin, 1999). From doing this work Allison and her team know that children like to:

- be in control and not to be controlled
- create things
- express themselves
- be social
- collaborate with other children

(a) What kind of tasks should be considered in evaluating a fluffy robot toy dog that can be programmed to move and to tell personalized stories about itself and children? The target age group for the toy is 7–9 years.

(b) Suggest heuristics to evaluate the toy.

Comment

(a) Tasks that you could consider: making the toy tell a story about the owner and two friends, making the toy move across the room, turn, and speak. You probably thought of others.

(b) The heuristics could be written to cover: being in control, being flexible, supporting expression, being motivating, supporting collaboration and being engaging. These are based on the issues raised by Druin, but the last one is aesthetic and tactile. Several of the heuristics needed would be more concerned with user experience (e.g., motivating, engaging, etc.) than with usability.

13.5 Asking experts: walkthroughs

Walkthroughs are an alternative approach to heuristic evaluation for predicting users' problems without doing user testing. As the name suggests, they involve walking through a task with the system and noting problematic usability features. Most walkthrough techniques do not involve users. Others, such as pluralistic walkthroughs, involve a team that includes users, developers, and usability specialists.

In this section we consider cognitive and pluralistic walkthroughs. Both were originally developed for desktop systems but can be applied to web-based systems, handheld devices, and products such as VCRs.

13.5.1 Cognitive walkthroughs

"Cognitive walkthroughs involve simulating a user's problem-solving process at each step in the human-computer dialog, checking to see if the user's goals and memory for actions can be assumed to lead to the next correct action." (Nielsen and Mack, 1994, p. 6). The defining feature is that they focus on evaluating designs for ease of learning—a focus that is motivated by observations that users learn by exploration (Wharton et al., 1994). The steps involved in cognitive walkthroughs are:

1. The characteristics of typical users are identified and documented and sample tasks are developed that focus on the aspects of the design to be evaluated. **A** description or prototype of the interface to be developed is also produced, along with a clear sequence of the actions needed for the users to complete the task.
2. A designer and one or more expert evaluators then come together to do the analysis.
3. The evaluators walk through the action sequences for each task, placing it within the context of a typical scenario, and as they do this they try to answer the following questions:
   - Will the correct action be sufficiently evident to the user? (Will the user know what to do to achieve the task?)
   - Will the user notice that the correct action is available? (Can users see the button or menu item that they should use for the next action? Is it apparent when it is needed?)

- Will the user associate and interpret the response from the action correctly? (Will users know from the feedback that they have made a correct or incorrect choice of action?)

In other words: will users know what to do, see how to do it, and understand from feedback whether the action was correct or not?

4. As the walkthrough is being done, a record of critical information is compiled in which:
   - The assumptions about what would cause problems and why are recorded. This involves explaining why users would face difficulties,
   - Notes about side issues and design changes are made.
   - A summary of the results is compiled.
5. The design is then revised to fix the problems presented.

It is important to document the cognitive walkthrough, keeping account of what works and what doesn't. **A** standardized feedback form can be used in which answers are recorded to the three bulleted questions in step **(3)** above. The form can also record the details outlined in points 1-4 as well as the date of the evaluation. Negative answers to any of the questions are carefully documented on a separate form, along with details of the system, its version number, the date of the evaluation, and the evaluators' names. It is also useful to document the severity of the problems, for example, how likely a problem is to occur and how serious it will be for users.

The strengths of this technique are that it focuses on users' problems in detail, yet users do not need to be present, nor is a working prototype necessary. However, it is very time-consuming and laborious to do. Furthermore the technique has a narrow focus that can be useful for certain types of system but not others.

Example: Find a book at Amazon.com

This example shows a cognitive walkthrough of buying this book at Amazon.com.

Task: to buy a copy of this book from Amazon.com
Typical users: students who use the web regularly

The steps to complete the task are given below. Note that the interface for Amazon.com may have changed since we did our evaluation.

Step 1. Selecting the correct category of goods on the home page

Q. Will users know what to do?

Answer: Yes—they know that they must find "books."

Q. Will users see how to do it?

Answer: Yes—they have seen menus before and will know to select the appropriate item and click go.

Q. Will users understand from feedback whether the action was correct or not?

Answer: Yes—their action takes them to a form that they need to complete to search for the book.

Step 2. Completing the form

Q. Will users know what to do?

Answer: Yes—the online form is like a paper form so they know they have to complete it.

Answer: No—they may not realize that the form has defaults to prevent inappropriate answers because this is different from a paper form.

Q. Will users see how to do it?

Answer: Yes—it is clear where the information goes and there is a button to tell the system to search for the book.

Q. Will users understand from feedback whether the action was correct or not?

Answer: Yes—they are taken to a picture of the book, a description, and purchase details.

ACTIVITY 13.10 Activity 13.7 was about doing a heuristic evaluation of REI.com or a similar e-commerce retail site. Now go back to that site and do a cognitive walkthrough to buy something, say a pair of skis. When you have completed the evaluation, compare your findings from the cognitive walkthrough technique with those from heuristic evaluation.

Comment You probably found that the cognitive walkthrough took longer than the heuristic evaluation for evaluating the same part of the site because it examines each step of a task. Consequently, you probably did not see as much of the website. It's likely that you also got much more detailed findings from the cognitive walkthrough. Cognitive walkthrough is a useful technique for examining a small part of a system in detail, whereas heuristic evaluation is useful for examining whole or parts of systems.

Variation of the cognitive walkthrough

A useful variation on this theme is provided by Rick Spencer of Microsoft, who adapted the cognitive walkthrough technique to make it more effective with a team who were developing an interactive development environment (IDE) (Spencer, 2000). When used in its original state, there were two major problems. First, answering the three questions in step (3) and discussing the answers took too long. Second, designers tended to be defensive, often invoking long explanations of cognitive theory to justify their designs. This second problem was particularly difficult because it undermined the efficacy of the technique and the social relationships of team members. In order to cope with these problems Rick Spencer adapted the technique by reducing the number of questions and curtailing discussion. This meant that the analysis was more coarse-grained but could be completed in much less time (about 2.5 hours). He also identified a leader, the usability specialist, and set strong ground rules for the session, including a ban on defending a design, debating cognitive theory, or doing designs on the fly.

These adaptations made the technique more usable, despite losing some of the detail from the analysis. Perhaps most important of all, he directed the social interactions of the design team so that they achieved their goal.

13.5.2 Pluralistic walkthroughs

"Pluralistic walkthroughs are another type of walkthrough in which users, developers and usability experts work together to step through a [task] scenario, discussing usability issues associated with dialog elements involved in the scenario steps" (Nielsen and Mack, **1994,** p. 5). Each group of experts is asked to assume the role of typical users. The walkthroughs are then done by following a sequence of steps (Bias, **1994):**

1. Scenarios are developed in the form of a series of hard-copy screens representing a single path through the interface. Often just two or a few screens are developed.
2. The scenarios are presented to the panel of evaluators and the panelists are asked to write down the sequence of actions they would take to move from one screen to another. They do this individually without conferring with one another.
3. When everyone has written down their actions, the panelists discuss the actions that they suggested for that round of the review. Usually, the representative users go first so that they are not influenced by the other panel members and are not deterred from speaking. Then the usability experts present their findings, and finally the developers offer their comments.
4. Then the panel moves on to the next round of screens. This process continues until all the scenarios have been evaluated.

The benefits of pluralistic walkthroughs include a strong focus on users' tasks. Performance data is produced and many designers like the apparent clarity of working with quantitative data. The approach also lends itself well to participatory design practices by involving a multidisciplinary team in which users play a key role. Limitations include having to get all the experts together at once and then proceed at the rate of the slowest. Furthermore, only a limited number of scenarios, and hence paths through the interface, can usually be explored because of time constraints.

Assignment

This assignment continues the work you did on the web-based ticketing system at the end of Chapters 7 and 8. The aim of this assignment is to evaluate the prototypes produced in the assignment of Chapter 8. The assignment takes an iterative form in which we ask you to evaluate and redesign your prototypes, following the iterative path in the interaction design process described in Chapter 6.

(a) For each prototype, return to the feedback you collected in Chapter 8 but this time perform open-ended interviews with a couple of potential users.

(b) Based on the feedback from this first evaluation, redesign the software/HTML prototype to take comments on all three prototypes into account.

(c) Decide on an appropriate set of heuristics and perform a heuristic evaluation of the redesigned prototype.

(d) Based on this evaluation, redesign the prototype to overcome the problems you encountered.

(e) Design a questionnaire to evaluate the system. The questionnaire may be paper-based or electronic. If it is electronic, make your software prototype and the questionnaire available to others and ask a selection of people to evaluate the system.

Summary

Techniques for asking users for their opinions vary from being unstructured and open-ended to tightly structured. The former enable exploration of concepts, while the latter provide structured information and can be replicated with large numbers of users, as in surveys. Predictive evaluation is done by experts who inspect the designs and offer their opinions. The value of these techniques is that they structure the evaluation process, which can in turn help to prevent problems from being overlooked. In practice, interviews and observations often go hand in hand, as part of a design process.

Key points

- There are three styles of interviews: structured, semi-structured and unstructured.
- Interview questions can be open or closed. Closed questions require the interviewee to select from a limited range of options. Open questions accept a free-range response.
- Many interviews are semi-structured. The evaluator has a predetermined agenda but will probe and follow interesting, relevant directions suggested by the interviewee. A few structured questions may also be included, for example to collect demographic information.
- Structured and semi-structured interviews are designed to be replicated.
- Focus groups are a form of group interview.
- Questionnaires are a comparatively low-cost, quick way of reaching large numbers of people.
- Various rating scales exist including selection boxes, Likert, and semantic scales.
- Inspections can be used for evaluating requirements, mockups, functional prototypes, or systems.
- Five experts typically find around 75% of the usability problems.
- Compared to user testing, heuristic evaluation is less expensive and more flexible.
- User testing and heuristic evaluation often reveal different usability problems.
- Other types of inspections include pluralistic and cognitive walkthroughs.
- Walkthroughs are very focused and so are suitable for evaluating small parts of systems.

Further reading

NIELSEN, J., AND MACK, R. L. (eds.) (1994) ***Usability Inspection Methods.*** New York: John Wiley & Sons. This book contains an edited collection of chapters on a variety of usability inspection methods. There is a detailed description of heuristic evaluation and walkthroughs and comparisons of these techniques with other evaluation techniques, particularly user testing. Jakob Nielsen's website *useit.com* provides additional information and advice on website design.

OPPENHEIM, A. N. (1992) ***Questionnaire Design, Interviewing and Attitude Measurement.*** London: Pinter Publishers. This text is useful for reference. It provides a detailed account of all aspects of questionnaire design, illustrated with many examples.

PREECE, J. ***(2000) Online Communities: Designing Usability, Supporting Sociability.*** Chichester, UK: John Wiley & Sons. This book is about the design of web-based online communities. It suggests guidelines for evaluating for sociability and usability that can be used as a basis for heuristics.

ROBSON, C. (1993) ***Real World Research.*** Blackwell. Oxford, UK. Chapter 9 provides basic practical guidance on how to interview and design questionnaires. It also contains many examples.

SHNEIDERMAN, B. (1998) ***Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd Edition)*** Reading, ***MA.:*** Addison-Wesley. Chapter 4 contains a discussion of the QUIS questionnaire.

INTERVIEW with Jakob Nielsen

Jakob Nielsen is a pioneer of heuristic evaluation. He is currently principal of the Nielsen Norman Consultancy Group and the author of numerous articles and books, including his recent book, *Designing Web Usability* (New Riders Publishing). He is well-known for his regular sound bites on usability which for many years have appeared at useit.com. In this interview Jakob talks about heuristic evaluation, why he developed the technique, and how it can be applied to the web.

JP: Jakob, why did you create heuristic evaluation?

JN: It is part of a larger mission I was on in the mid-'80s, which was to simplify usability engineering, to get more people using what I call "discount usability engineering." The idea was to come up with several simplified methods that would be very easy and fast to use. Heuristic evaluation can be used for any design project or any stage in the design process, without budgetary constraints. To succeed it had to be fast, cheap, and useful.

JP: How can it be adapted for the web?

JN: I think it applies just as much to the web, actually if anything more, because a typical website will have tens of thousands of pages. A big one may have hundreds of thousands of pages, much too much to be assessed using traditional usability evaluation methods such as user testing. User testing is good for testing the home page or the main navigation system. But if you look at the individual pages, there is no way that you can really test them. Even with the discount approach, which would involve five users, it would still be hard to test all the pages. So all you are left with is the notion of doing a heuristic evaluation, where you just have a few people look at the majority of pages and judge them according to the heuristics. Now the heuristics are somewhat different, because people behave differently on the web. They are more ruthless about getting a very quick glance at what is on a page and if they don't understand it then leaving it. Typically application users work a little harder at learning an application. The basic heuristics that I developed a long time ago are universal, so they apply to the web as well. But as well as these global heuristics that are always true, for example "consistency," there can be specialized heuristics that apply to particular systems. But most evaluators use the general heuristics because the web is still evolving and we are still in the process of determining what the web-specific heuristics should be.

JP: So how do you advise designers to go about evaluating a really large website?

JN: Well, you cannot actually test every page. Also, there is another problem: developing a large website is incredibly collaborative and involves a lot of different people. There may be a central team in charge of things like the home page, the overall appearance, and the overall navigation system. But when it comes to making a product page, it is the product-marketing manager of, say, Kentucky who is in charge of that. The division in Kentucky knows about the product line and the people back at headquarters have no clue about the details. That's why they have to do their own evaluations in that department. The big thing right now is that this is not being done, developers are not evaluating enough. That's one of the reasons I want to push the heuristic evaluation method even further to get it out to all the website contributors. The uptake of usability methods has dramatically improved from five years ago, when many companies didn't have a clue, but the need today is still great because of the phenomenal development of the web.

JP: When should you start doing heuristic evaluation?

JN: You should start quite early, maybe not quite as early as testing a very rough mockup, but as soon as there is a slightly more substantial prototype. For example, if you are building a website that might eventually have ten thousand pages, it would be appropriate to do a heuristic evaluation of, say, the first ten to twenty pages. By doing this you would catch quite a lot of usability problems.

JP: How do you combine user testing and heuristic evaluation?

JN: I suggest a sandwich model where you layer them on top of each other. Do some early user testing of two or three drawings. Develop the ideas somewhat, then do a heuristic evaluation. Then evolve the design further, do some user tests, evolve it and do heuristic evaluation, and so on. When the design is nearing completion, heuristic evaluation is very useful particularly for a very large design.

JP: So, do you have a story to tell us about your consulting experiences, something that opened your eyes or amused you?

JN: Well, my most interesting project started when I received an email from a co-founder of a large company who wanted my opinion on a new idea. We met and he explained his idea and because I know a lot about usability, including research studies, I could warn him that it wouldn't work—it was doomed. This was very satisfying and seems like the true role for a usability consultant. I **think** usability consultants should have this level of insight. It is not enough to just clean up after somebody makes the mistake of starting the wrong project or produces a poor design. We really should help define which projects should be done in the first place. Our role is to help identify options for really improving people's lives, for developing products that are considerably more efficient, easier or faster to learn, or whatever the criteria are. That is the ultimate goal of our entire field.

JP: One last question—how do you think the web will develop? What will we see next, what do you expect the future to bring?

JN: I hope we will abandon the page metaphor and reach back to the earlier days of hypertext. There are other ideas that would help people navigate the web better. The web is really an "article-reading" interface. My website useit.com, for example, is mainly articles, but for many other things people need a different interface, the current interface just does not work. I hope we will evolve a more interesting, useful interface that I'll call the "Internet desktop," which would have a control panel for your own environment, or another metaphor would be "your personal secretary." Instead of the old goal where the computer spits out more information, the goal would be for the computer to protect you from too much information. You shouldn't have to actually go and read all those webpages. You should have something that would help you prioritize your time so you would get the most out of the web. But, pragmatically speaking, these are not going to come any time soon. My prediction has been that Explorer Version 8 will be the first good web browser and that is still my prediction, but there are still a few versions to come before we reach that level. The more short-term prediction is really that designers will take much more responsibility for content and usability of the web. We need to write webpages so that people can read them. For instance, we need headlines that make sense. Even something as simple as a headline is a user interface, because it's now being used interactively, not as in a magazine where you just look at it. So writing the headline, writing the content, designing the navigation are jobs for the individual website designers. In combination, such decisions are really defining the user experience of the network economy. That's why we really have an obligation, every one of us, because we are building the new world and if the new world turns out to be miserable, we have only ourselves to blame, not Bill Gates. We've got to design the web for the way users behave.