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 HOME RUN:

• **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.](image-url)
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).

The text in this document is from the book: