A Practical Guide to Usability Testing

REVISED EDITION

Joseph S. Dumas
Janice C. Redish

intellect

9

Deciding Who Should Be Participants

One of the cardinal rules of usability testing is that the people who work with the product in the usability test must be like the people who will actually use the product.

The participants must represent the people who will use the product.

If the product is for office workers, software engineers are not appropriate participants. If the product is meant for experienced oscilloscope operators, students who are just learning oscilloscopes are not appropriate.

As you decide who to have as participants, you'll go through these steps, which we cover in the rest of this chapter:

- developing user profiles
- selecting subgroups for a test
- defining and quantifying characteristics for each subgroup
- deciding how many people to include in a test

Developing User Profiles

Ideally, of course, you or others will have developed a profile of the actual and potential users of the product, long before the time for a usability test. If it hasn't been done, however, you have to do it as one of the first steps in planning the test.

To get useful results from a usability test, you must know the users—and potential users—of the product.

Developing a good profile of users should be a joint effort of the marketing department, usability specialists, and product designers. If this is a new product, the user profile should be based on general market research, on an analysis of the customers of competitors' products, on focus group sessions, and on observing and interviewing prospective users. If you will be testing enhancements to an existing product, the user profile should be based on surveys, focus group sessions, contextual interviews, and usability tests of the current version of the product. These are some of the techniques that we discussed in Chapter 3 in the first part of the book.

Unfortunately, products are sometimes still being designed without a clear understanding of the needs and skills of the people who will use them. In that case, you may have to prepare profiles of the users as you plan the usability test.

Thinking About Relevant Characteristics

In developing a profile of users, you want to capture two types of characteristics:

- those that all users will share
- those that might make a difference among the users

Consider the electronic mail program we are using as our ongoing example. This product is aimed at the corporate, rather than the

home, market. The software company that is developing this product has decided that the market they want to aim for is mid- to large-size corporations. When we look for participants for the usability test, therefore, we will want to look for people who work in mid- to large-size corporations and who have reason to communicate often with others in the company.

Some of the companies that are potential customers will be introducing electronic mail to their employees for the first time. Others will be switching employees from another electronic mail product to this one. Therefore, in the usability test, we will want to include some participants who are new to electronic mail and some who have experience with other electronic mail programs.

Other key decisions we have to make about this test are

- whether to include participants who have no experience with computers at all
- whether to include participants who have used only stand-alone personal computers and who have never logged into a network
- whether to include participants who have used terminals connected to a mainframe, but who have never used a menu-driven, personal computer application like this one

The appropriate decisions depend, of course, on the market for the product. We would have to ask for more information about the market.

> In organizations that will be buying this as their first electronic mail product, will many users also be getting on computers for the first time? Will many users be experienced at personal computer applications but be new to the networking environment? Will many users already be adept at other menu-driven products?

When you probe for information like this, don't be surprised if you get fuzzy answers like "It depends" or "All of the above." Keep pressing to find the people who know the market and then to find out the answers to your questions. If, as indeed may be the case, the market includes people in all of these categories, the team has to decide which of these users are your primary concern for this usability test. If this is your first usability test for the product, you may want to focus on "typical" users—those who represent the largest part of the market. If you have already tested with "typical" users, you may have a particular concern for users who represent a smaller part of the market.

Deciding Which Factors Matter Most

In our work, we have found that specific, relevant experience and motivation matter more for understanding differences in how people

interact with products than do demographic factors like education, income level, and age. Most products that we test are designed for people with average physical and mental abilities. Of course, if the product is meant for a population with special characteristics, such as an educational program for children, a service for the elderly, or a teletype machine for hearing-impaired people, you would make the relevant demographic characteristics an important part of the user profile.

Here are some of the factors to consider in developing a profile of a product's potential users:

- work experience, such as:
  - job description
  - length of experience on the job
  - length of experience with the particular tasks that the product handles
- general computer experience, such as:
  - length of experience working with a computer
  - types of applications used
  - length of time using each application
  - frequency of using each application
- specific computer experience, such as:
  - length of experience with specific, relevant hardware (mouse, function keys, pen-based, etc.)
  - frequency of experience with that hardware
  - length of experience with the relevant operating system (UNIX, Windows, Macintosh, etc.)
- experience with this product, such as:
  - length and frequency of using basic features
  - length and frequency of using advanced features
- experience with similar products, such as:
  - length and frequency of using basic features
  - length and frequency of using advanced features

Thinking Broadly About Users

As you develop profiles of the users, we urge you to *think broadly*. Here are four examples of what we mean by thinking broadly about users:

1. Don't just consider people who are now doing this job. Think about job turnover and new hires. In most companies, although a job description remains the same for many years, the people who actually do that job change frequently. What's the rate of turnover in the job that this product is meant to handle?
2. Don't limit yourself to the current market. Growth for most companies means getting more people to accept and use the

product. Who are these new people, in terms of the characteristics that you have said might make a difference in how easy the product is to use or learn?

3. If you are developing an internal product, don't necessarily limit yourself to the group for which the product is being developed. Think about future expansion. What other groups in the company might be using this product a year from now? Would it ever become an external product, and, if so, who would be the market for it?
4. Think about differences within a category. For example, experienced workers and novice workers may have very different reactions to a new product.
   On the one hand, if the product is structured by the tasks of the job or is similar to something they have used before, experienced workers may find the product easy to use, while novice workers may find it difficult.
   On the other hand, if the product is introducing a new interface or is changing the way that the job has been done, experienced workers may have a more difficult time with it than novices. If the novice workers are also younger, they may have more experience with computers in general and less fear or reluctance to try new methods.

As you plan the usability test, plan to include participants from both ends of the range of each category that you choose to include.

At the end of this chapter, you'll find a form called "Developing a user profile." You'll also find two filled-out examples of the form with explanations.

When you understand who the users are and have thought about which characteristics might make a difference in the usability test results, you have to decide how to group people for the test and which subgroups of users to include in the test. That's the topic of the next section.

Selecting Subgroups for a Test

A subgroup is composed of the people who share specific characteristics that are important in the user profile. For our electronic mail example, one subgroup might be people who have previous experience using electronic mail.

Dividing Users into Subgroups

As you add and combine characteristics, the number of subgroups that you have grows very quickly. To understand how this works, let's consider a different example.

Suppose the product you are going to test is a billing system for law offices. The billing system will be used by both lawyers and their secretaries, and you believe they may have different reactions to the product. At this point, you have two subgroups:

1. lawyers
2. legal secretaries

You believe the product's acceptance and ease of use is going to depend on how comfortable the person already is with computers. Adding that characteristic at two levels of experience gives you four subgroups:

1. lawyers with "a lot of" computer experience
2. lawyers with "very little" computer experience
3. legal secretaries with "a lot of" computer experience
4. legal secretaries with "very little" computer experience

You will have to define what you mean by "a lot of" computer experience and what you mean by "very little" computer experience. We'll come back in a moment to the problem of defining levels within a particular characteristic.

Adding Another Characteristic

Let's continue thinking about the relationship between different characteristics and the number of subgroups. Suppose that you also decide that previous experience doing billing might make a difference in the product's acceptance and ease of use. If you were to add that characteristic to the usability test, you would have eight subgroups to test:

Lawyers with:

1. a lot of computer experience and also experience in billing
2. a lot of computer experience but no experience in billing
3. very little computer experience but with experience in billing
4. very little computer experience and no experience in billing

Legal secretaries with:

5. a lot of computer experience and also experience in billing
6. a lot of computer experience but no experience in billing
7. very little computer experience but with experience in billing
8. very little computer experience and no experience in billing

You can see that the number of subgroups increases exponentially as you add each new characteristic. Adding a fourth characteristic with two levels would give you 16 subgroups.

Selecting the Most Critical Characteristics

You are likely to have the time and money to include only two to four subgroups in a test. You usually have to select the one or two characteristics that are most relevant to your goals and concerns.

Some subgroups may be less critical to your goals and concerns than others. For the new billing system, you might be most concerned about lawyers who are new to computers. If budget or time constraints forced you to have fewer subgroups than you would like, you might decide to make a painful tradeoff and not include experienced computer users in your test, even though you realize that usability problems that affect only experienced computer users may not show up in the test. Having a usability expert review the product from the experienced computer user's point of view might make the tradeoff less painful.

Gathering Other Information

You can—and should—gather information about the other characteristics. If you find anomalies in the test data, you can look back at your information on each test participant and see if one of the other characteristics helps to explain the anomalies.

> Let's say that you decide to run the usability test of the legal billing software with groups divided by job description and computer experience. When you analyze the results, you find that one of the computer-experienced lawyers found the product much easier to use than the others in that subgroup. You look back at the information you gathered on each test participant and find that the other lawyers had never done their own billing, but this lawyer had been billing her clients herself for many years. That experience in billing might explain the differences in the behavior that you observed.

You'll have that other information about each participant if you ask for it either on a survey form that you use as you recruit or on a pretest questionnaire that you give each participant when he or she arrives for the usability test. (You'll find an example of a survey form for recruiting in the next chapter on "Recruiting Participants." You'll find an example of a pretest questionnaire in Chapter 14 on "Preparing Test Materials.")

Defining and Quantifying Characteristics of Each Subgroup

To decide who fits into your subgroups, you have to define and quantify each characteristic for each subgroup. In our example of the billing system, you have to decide what to count as "very little computer experience" and as "a lot of computer experience."

You might define *computer experience* to mean "uses a personal computer with business software" and "very little" as "none to three months worth" and "a lot of" as "at least 1 year." This distinction would be based on your belief from interviews and observations that someone who has used a computer for only up to three months is still learning about using it, while someone who has been using a computer for a year or more understands most of the keyboard and screen conventions. You could also define computer experience by the number of different applications used or by frequency of use rather than by time, if either of those criteria is a better indicator of experience for users of your product.

Leaving Out "Intermediate" Users

Note that these definitions of "very little" and "a lot of" leave out a middle group. People who have used the computer regularly for more than 3 months but less than 1 year would not fit into either category. You would not select them as participants in this test.

For many categories, you could set three or more levels of experience. For "length of time using a computer," you could say that relevant subgroups would be:

novice	0 to 3 months
intermediate	more than 3 months, less than 1 year
experienced	more than 1 year

Selecting participants from the extremes of the range for a particular characteristic will often give you more useful information than selecting from the middle.

If you cannot include 3 to 5 people from each of these subgroups in your test, however, we suggest that you drop the middle group.

If you find that participants from both the novice and experienced subgroups have the same problem, you will feel comfortable generalizing that intermediate users are also likely to have that problem.

Selecting a Range of Participants Within Each Subgroup

Even within each subgroup, you will find a range of potential participants. If you set a requirement of at least 1 year of personal computer experience, you may well find participants who have much more experience than that. Try to get a range of people within the subgroup. If all of your participants have close to 1 year of experience, you may be missing what will happen when the product gets to people who have 5 years of experience.

Setting a Minimum and Maximum for a Subgroup

For some tests, you may want to exclude people at the extreme of a subgroup. You might want to set both a minimum and a maximum requirement for a particular subgroup.

Let's say that your concern is how easily the Windows version of a spreadsheet will be for people who have been using that program in a DOS environment. You are particularly concerned about "typical," "average" users, not about "hackers" or "power" users. One of your subgroups might be people who have used the program at least 3 times a week for 6 months or more in a DOS environment. That would be your minimum requirement. To define and separate the "typical" or "average" user from the "hacker" or "power" user, you might say that you will exclude participants who have *created*, not just used, macros in this spreadsheet program.

Deciding How Many People to Include in a Test

The issue of how many participants you need in a usability test is one of intense interest and discussion in the usability community. Remember that a usability test is not a research study. The purpose of a usability test is to uncover the most serious problems that users are likely to have with a product.

Nielsen and Molich (1990) found that not quite half of all major usability problems were detected with three participants. Virzi (1992) found that 80% of the usability problems in a product were detected with between 4 and 5 participants and 90% were detected with 10 participants. In addition, all of the *global* problems in the product he tested were detected with 10 participants. Additional participants were less and less likely to reveal new information.

The number of people to include as participants in a test depends on

- how many subgroups you need to satisfy your goals and concerns
- how much time and money you can get for the test
- how important it is for you to be able to compute statistical significance for your results

Most of the time, you will not want to use inferential statistics, that is to measure results to a particular level of confidence and range of error in a statistical test. However, if an entire product line hinges on what you are testing, or if the product represents a particularly large investment, the company may want to include enough participants for you to use statistical tests of significance. (See Chapter 20, "Tabulating and Analyzing Data," for more on the place of inferential statistics in usability testing.)

A Typical Test Includes 6 to 12 Participants

A typical usability test now includes 6 to 12 participants in two to three subgroups. As you can see from Nielsen and Molich's and Virzi's results, you need 3 to 5 participants in each subgroup to feel comfortable that you are seeing the problems. Three participants for each subgroup is probably an absolute minimum. With fewer people in a subgroup, you won't know if you are seeing idiosyncratic behavior or something that is likely to generalize across the subgroup. Even if you are testing an early prototype, you should still plan on at least 3 people representing each type of user.

You See the Same Problem Many Times

With 3 to 5 people in each subgroup, you are likely to see enough to feel comfortable with the conclusions that you reach. After you've seen several people make the same mistake, you don't need to see it a 10th or 20th or 50th time. You'll know that you've uncovered a problem. Moreover, when an interface has global problems, you will see the same problem many times with each participant.

> If a menu hierarchy is poorly structured, you are likely to see each participant fail several times to find the correct option. Although the option the participant is looking for may differ from task to task, the problem will be the same—the user's assumption about where to find the option does not match the hierarchy that the designers built.

You may see some problems that are common to and confined to one of your subgroups. After all, you divide users into these subgroups for testing *because* you are concerned that they will have different problems or that one subgroup will have a problem where another does not. However, you are likely to also see many problems that cut across your subgroups. In many situations, your quantitative data will show that all or almost all the participants, regardless of their background, had the same problem.

You Have to Balance Time, Money, and Information Gained

Time and money, unfortunately, most often determine the number of participants in a usability test. Because most usability tests are conducted 1 participant at a time, and a typical test session lasts half a day, a test with 10 participants takes a week of lab time. More participants obviously means more time in the lab. Having more participants may also increase the time you need for recruiting and the time you need to analyze the results.

Some teams do shorter tests, particularly when they are testing

prototypes. They have 3 to 6 people a day, for 1 to 2 hours each. That way, they can complete the observing and recording part of the test in a shorter time or watch more people working with the product. Be careful, however, not to overwhelm the test team, particularly if they are new to usability testing. Being in the laboratory with the participants is very intensive, often very tiring, work.

The determining factors in how many people to include in a usability test should be not only time and money, but also the number of subgroups that you need in order to cover all the important characteristics that differentiate groups of potential buyers and users.

Deciding how many participants to include is always a balancing act. On the one hand, you want results quickly and at minimal cost. On the other hand, you want to be sure that what you are seeing is typical of what will happen to the product when it is released. To make the most of the limited number of participants in a typical usability test, you must

- decide carefully which characteristics are the most important so that you define subgroups that will be most useful
- collect other relevant information to help you account for other differences that show up in the results
- select people for each subgroup who are representative of the full range of qualifications for that subgroup

In this chapter, we've looked in detail at the first two of these items. We'll cover more about the last item in the next chapter on "Recruiting Participants."

Examples of Forms to Help in Developing User Profiles

The following pages contain examples of:

- a blank form for developing a user profile for a usability test (Figure 9-1)
- two filled out examples of the form (Figures 9-2 and 9-3)
- an explanation of each example

Developing a user profile for a usability test

1. **Product name:** _____

2. **General characterization of the user population:**

   _____

   _____

3. **Characteristics of the users that are relevant to the test:**

   _____

   _____

   _____

   _____

4. **Which of the characteristics that you listed in 3. should all users in the test have in common and how will you define them?**

   _____

   _____

   _____

5. **Which of the characteristics that you listed in 3. will vary in the test and how will you define them?**

   _____

   _____

   _____

Figure 9-1. Sample form for user profiles for a usability test

Developing a user profile

1. **Product name:** `New billing software`

2. **General characterization of the user population:**

   ```
   lawyers and

   their secretaries or office managers
   ```

3. **Characteristics of the users that are relevant to the test:**

   ```
   computer experience
   job experience
   billing experience
   other database experience
   ```

4. **Which of the characteristics that you listed in 3. should all users in the test have in common and how will you define them?**

   ```
   lawyer in private practice for at
   least one year

   legal secretary or office manager for
   at least one year for lawyers in
   private practice
   ```

5. **Which of the characteristics that you listed in 3. will vary in the test and how will you define them?**

   ```
   computer experience

   "very little" = 0 to 3 months

   "a lot of" = 12 months or more
   ```

Figure 9-2. A user profile for a test of Legal Sofware's billing system

Explanation of the Example in Figure 9-2

Legal Software is developing a new billing system for law offices. They come to The Testing Laboratory (TTL) to find out how usable their new product is. Working with the designers from Legal Software, usability test specialists from TTL identify the market for Legal Billing as all lawyers who bill clients and their legal secretaries or office managers (Question 2).

Thinking about the way that people will use Legal Billing, TTL specialists and the client decide that acceptance and ease of use of the

product may depend on factors like other computer experience, whether they already use a billing system, experience with other software of the same genre (databases), and how much experience they have at their jobs (Question 3).

Because Legal Software is particularly concerned about whether the product will be better accepted and used by lawyers or their secretaries/office managers, the team decides to include a group of each and then defines the characteristics that will put a person in one or the other of those groups (Question 4).

Legal Software wants their product to be useful even to people who may buy a computer just to use this product. Therefore, they are very much concerned about how much prior computer experience is needed to make these users comfortable with the product. The test team decides to form subgroups based on computer experience (Question 5).

Legal Software is also concerned about whether people who don't now handle their own billing will find their product so easy to learn and use that they will switch to it. However, they cannot afford to have more subgroups, so TTL agrees to ask questions about experience in billing in a pretest questionnaire. They'll also ask questions about the other factors that they identified in Question 3.

Explanation of the Example in Figure 9-3

You are developing our example electronic mail software. You are especially concerned about introducing this software in mid- to large-size companies and being sure that decision makers who have to communicate will find it easy to use. You want to capture both the market of users who are new to electronic mail and those whom you can convince to switch from another product to yours (Question 2).

You decide that the characteristics of users that are relevant to this usability test are

- size of the company (number of employees)
- job level of participants (managers or not)
- experience using electronic mail software
- computer experience
- experience using applications in a networking environment (Question 3)

To focus on decision makers in the companies that are your primary market, you will include only people who work in organizations of at least 50 people and who manage at least 2 other people (Question 4).

The resources that you have for the test will only cover 10 participants. You can only have two subgroups of 5 participants each. You will have a group of "novices" defined as "managers who have

Developing a user profile

1. **Product name:** `E-mail`

2. **General characterization of the user population:**

```
employees in mid- to large-size
companies with local area networks
```

3. **Characteristics of the users that are relevant to the test:**

```
size of company
role in company (managers)
experience with electronic mail
experience with computers
experience with networks
```

4. **Which of the characteristics that you listed in 3. should all users in the test have in common and how will you define them?**

```
size of company:  50 or more people

role:  manager of at least two people
```

5. **Which of the characteristics that you listed in 3. will vary in the test and how will you define them?**

```
electronic mail experience
     "novice" = none
     "experienced" = used 6 months or
     more

networking experience
     "novice" = none
     "experienced" = used 6 months or
     more
```

Figure 9-3. A user profile for a test of our electronic mail example

never used an electronic mail program and who have no experience using a computer in a networking environment." You will have a group of "experienced users" defined as "managers who have used a different electronic mail program in a networking environment for at least 6 months" (Question 5).