



Handbook of Usability Testing

Second Edition

How to Plan, Design, and Conduct Effective Tests

Jeff Rubin
Dana Chisnell



Wiley Publishing, Inc.

CHAPTER 2

What Is Usability Testing?

The term *usability testing* is often used rather indiscriminately to refer to *any* technique used to evaluate a product or system. Many times it is obvious that the speaker is referring to one of the other techniques discussed in Chapter 1. Throughout this book we use the term *usability testing* to refer to a process that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria. This inclusion of representative users eliminates labeling as usability testing such techniques as expert evaluations, walk-throughs, and the like that do not require representative users as part of the process.

Usability testing is a research tool, with its roots in classical experimental methodology. The range of tests one can conduct is considerable, from true classical experiments with large sample sizes and complex test designs to very informal qualitative studies with only a single participant. Each testing approach has different objectives, as well as different time and resource requirements. The emphasis of this book is on more informal, less complex tests designed for quick turnaround of results in industrial product development environments.

Why Test? Goals of Testing

From the point of view of some companies, usability testing is part of a larger effort to improve the profitability of products. There are many aspects to doing so, which in the end also benefits users greatly: design decisions are informed by data gathered from representative users to expose design issues so they can be remedied, thus minimizing or eliminating frustration for users.

Informing Design

The overall goal of usability testing is to inform design by gathering data from which to identify and rectify usability deficiencies existing in products and their accompanying support materials prior to release. The intent is to ensure the creation of products that:

- Are useful to and valued by the target audience
- Are easy to learn
- Help people be effective and efficient at what they want to do
- Are satisfying (and possibly even delightful) to use

Eliminating Design Problems and Frustration

One side of the profitability coin is the ease with which customers can use the product. When you minimize the frustration of using a product for your target audience by remedying flaws in the design ahead of product release, you also accomplish these goals:

- Set the stage for a positive relationship between your organization and your customers.
- Establish the expectation that the products your organization sells are high quality and easy to use.
- Demonstrate that the organization considers the goals and priorities of its customers to be important.
- Release a product that customers find useful, effective, efficient, and satisfying.

Improving Profitability

Goals or benefits of testing for your organization are:

- **Creating a historical record of usability benchmarks for future releases.** By keeping track of test results, a company can ensure that future products either improve on or at least maintain current usability standards.
- **Minimizing the cost of service and support calls.** A more usable product will require fewer service calls and less support from the company.
- **Increasing sales and the probability of repeat sales.** Usable products create happy customers who talk to other potential buyers or users. Happy customers also tend to stick with future releases of the product, rather than purchase a competitor's product.

- **Acquiring a competitive edge because usability has become a market separator for products.** Usability has become one of the main ways to separate one's product from a competitor's product in the customer's mind. One need only scan the latest advertising to see products described using phrases such as "simple" and "easy" among others. Unfortunately, this information is rarely truthful when put to the test.
- **Minimizing risk.** Actually, all companies and organizations have conducted usability testing for years. Unfortunately, the true name for this type of testing has been "product release," and the "testing" involved trying the product in the marketplace. Obviously, this is a very risky strategy, and usability testing conducted *prior* to release can minimize the considerable risk of releasing a product with serious usability problems.

Basics of the Methodology

The basic methodology for conducting a usability test has its origin in the classical approach for conducting a controlled experiment. With this formal approach, often employed to conduct basic research, a specific hypothesis is formulated and then tested by isolating and manipulating variables under controlled conditions. Cause-and-effect relationships are then carefully examined, often through the use of the appropriate inferential statistical technique(s), and the hypothesis is either confirmed or rejected. Employing a true experimental design, these studies require that:

- **A hypothesis must be formulated.** A hypothesis states what you expect to occur when testing. For example, "Help as designed in format A will improve the speed and error rate of experienced users more than help as designed in format B." It is essential that the hypothesis be as specific as possible.
- **Randomly chosen (using a very systematic method) participants must be assigned to experimental conditions.** One needs to understand the characteristics of the target population, and from that larger population select a representative random sample. Random sampling is often difficult, especially when choosing from a population of existing customers.
- **Tight controls must be employed.** Experimental controls are crucial or else the validity of the results can be called into question, regardless of whether statistical significance is the goal. All participants should have nearly the identical experience as each other prior to and during the test.

In addition, the amount of interaction with the test moderator must be controlled.

- **Control groups must be employed.** In order to validate results, a control group must be employed; its treatment should vary only on the single variable being tested.
- **The sample (of users) must be of sufficient size to measure statistically significant differences between groups.** In order to measure differences between groups statistically, a large enough sample size must be used. Too small a sample can lead to erroneous conclusions.

The preceding approach is the basis for conducting classical experiments, and when conducting basic research, it is the method of choice. *However, it is not the method expounded in this book for the following reasons.*

- It is often impossible or inappropriate to use such a methodology to conduct usability tests in the fast-paced, highly pressurized development environment in which most readers will find themselves. It is impossible because of the many organizational constraints, political and otherwise. It is inappropriate because the purpose of usability testing is not necessarily to formulate and test specific hypotheses, that is, conduct research, but rather to make informed decisions about design to improve products.
- The amount of prerequisite knowledge of experimental method and statistics required in order to perform these kinds of studies properly is considerable and better left to an experienced usability or human factors specialist. Should one attempt to conduct this type of tight research without the appropriate background and training, the results can often be very misleading, and lead to a worse situation than if no research had been conducted.
- In the environment in which testing most often takes place, it is often very difficult to apply the principle of randomly assigning participants because one often has little control over this factor. This is especially true as it concerns the use of existing customers as participants.
- Still another reason for a less formal approach concerns sample size. To achieve generalizable results for a given target population, one's sample size is dependent on knowledge of certain information about that population, which is often lacking (and sometimes the precise reason for the test). Lacking such information, one may need to test 10 to 12 participants *per condition* to be on the safe side, a factor that might require one to test 40 or more participants to ensure statistically significant results.

- Last, and probably most important, the classical methodology is designed to obtain quantitative proof of research hypotheses that one design is better than another, for example. It is not designed to obtain qualitative information on how to fix problems and redesign products. We assume that most readers will be more concerned with the latter than the former.

The approach we advocate is a more informal, iterative approach to testing, albeit with experimental rigor at its core. As the reader will see in later chapters of this book, experimental rigor is essential for *any* study that one conducts.

Much can be achieved by conducting *a series* of quick, pointed studies, beginning early in the development cycle. It is the intent of this book to present the basics of conducting this type of less formal, yet well-designed test that will identify the specific usability deficiencies of a product, their cause, and the means to overcome them. The basics of this approach are described in the sections that follow.

Basic Elements of Usability Testing

- Development of research questions or test objectives rather than hypotheses.
- Use of a representative sample of end users which may or may not be randomly chosen.
- Representation of the actual work environment.
- Observation of end users who either use or review a representation of the product.
- Controlled and sometimes extensive interviewing and probing of the participants by the test moderator.
- Collection of quantitative and qualitative performance and preference measures.
- Recommendation of improvements to the design of the product.

We detail the “how-to” of this approach in the chapters that follow.

Limitations of Testing

Now, having painted a rather glorified picture of what usability testing is intended to accomplish, let’s splash a bit of cold water on the situation. Testing is neither the end-all nor be-all for usability and product success, and it is

important to understand its limitations. Testing does not guarantee success or even prove that a product will be usable. Even the most rigorously conducted formal test cannot, with 100 percent certainty, ensure that a product will be usable when released. Here are some reasons why:

- **Testing is always an artificial situation.** Testing in the lab, or even testing in the field, still represents a depiction of the actual situation of usage and not the situation itself. The very act of conducting a study can itself affect the results.
- **Test results do not prove that a product works.** Even if one conducts the type of test that acquires statistically significant results, this still does not prove that a product works. Statistical significance is simply a measure of the probability that one's results were not due to chance. It is not a guarantee, and it is very dependent upon the way in which the test was conducted.
- **Participants are rarely fully representative of the target population.** Participants are only as representative as your ability to understand and classify your target audience. Market research is not an infallible science, and the actual end user is often hard to identify and describe.
- **Testing is not always the best technique to use.** There are many techniques intended to evaluate and improve products, as discussed in Chapter 1 and Chapter 13. For example, in some cases it is more effective both in terms of cost, time, and accuracy to conduct an expert or heuristic evaluation of a product rather than test it. This is especially true in the early stages of a product when gross violations of usability principles abound. It is simply unnecessary to bring in many participants to reveal the obvious.

However, in spite of these limitations, usability testing, when conducted with care and precision, for the appropriate reasons, at the appropriate time in the product development lifecycle, and as part of an overall user-centered design approach, is an almost infallible indicator of potential problems and the means to resolve them. It minimizes the risk considerably of releasing an unstable or unlearnable product. In almost every case, and this is an underlying theme of this book: *it is better to test than not to test*.

The next chapter covers the basics for conducting four types of specific tests and then provides a hypothetical case study employing all four tests in the course of a development cycle.