

## Chapter 8

---

---

# Tools, Templates, and Testing Facilities

- If you throw usability staff into the organization without the right equipment, they are going to seem slow, inefficient, and impractical.
- Get tools (e.g., lab equipment), templates (e.g., reusable questionnaires), and testing facilities. These items form an essential toolkit—the core infrastructure for routine usability work.
- Your toolkit makes it efficient to complete the methodology. To determine the toolkit you need, review your methodology.

---

A well-trained staff in a room with nothing but paper can outdesign a poorly trained staff equipped with a state-of-the-art facility.

The main value of facilities, tools, and templates is time savings. Instead of creating a testing form from scratch every time a test is needed, a usability engineer can take an existing form and modify it for a client's specific test in about 20 minutes. Creating the concept

for a test and the forms from scratch takes days or even weeks. So, hire good staff members, and supply them with the tools that make a difference. This chapter outlines the tools you need, the templates that are helpful, and usability testing facilities that will help your staff be most efficient and effective.

Note, however, that by the time this book is published, some of the tools and templates described here may be outdated because new developments happen all the time. For example, you may hear that usability testing labs have recently moved from being “marginally useful in special circumstances” to becoming a practical part of almost every test. Or you may learn that remote testing, which isn’t used often today, is becoming far more practical and therefore much more widely used. **Remote testing** is usability testing performed at a distance; the participant and the facilitator will not be in the room together (in fact, may not be on the same continent), yet the facilitator can still monitor what the participant is doing and saying. Because toolsets will likely change, a skeptical attitude about these tools is useful—if a tool does not really make a difference in the design, spend your money another way.

---

## Introduction to Your Toolkit

Your methodology points to the facilities, tools, and templates you need. For example, if the methodology specifies that a test of branding occurs at a certain point, you will want to have templates for reusable questionnaires and a standard template for the final report.

If you update your methodology, you may need to update the corresponding tools, templates, and facilities. Also, new facilities, tools, and templates might lead you to change your methodology. For example, online prototyping has become easier, so you might move it further up into the design cycle. Or, as remote testing becomes more feasible and useful, you may add it to your methodology and develop new tools and templates to fit it. However, be careful about implementing these kinds of changes because some “amazing” breakthroughs are actually not that useful.

The following sections cover the infrastructure you should consider implementing at your company. They also explore scenarios and priorities for each facility.

---

## Testing Facilities

Depending on circumstances, testing facilities can range from a simple office setting or a hotel room to a full-blown usability testing lab. You do not have to have a full usability testing lab in order to conduct usability testing. If office space is at a premium, the office of one of the usability team members can be used for testing. There may not be a one-way mirror, special equipment, or videotaping. There may be only a few chairs, a desk, and a computer. However, skilled staff members can still successfully create and run the tests. Similarly, it is quite acceptable to use a conference room to run tests; however, it is critical that the room be reasonably quiet and free of visual and auditory interruptions. For this reason, it is best never to use participants' workspaces for testing. You can observe them there, but workspaces are not good places to run tests.

There are a number of reasons for having a formal and dedicated usability testing facility. One reason is that designating a space for testing shows a commitment to testing within the organization. It is nice to have a room or perhaps a suite with that label, but this will not have value so much in supporting the work as in making a political statement. Of course, the facility becomes an albatross if it is not regularly used. Unfortunately, some labs left unused later become storage spaces.

There is a real value to having a quality testing environment. While the results of running tests in storage closets can still be quite good, it is best to have a testing environment that makes the participants and the facilitator feel comfortable and important. If you can make the test a relaxed experience, you will get more accurate and complete results. At the same time, facilities that feel imposing and overly scientific should be avoided—you do not want the evaluation environment to feel too formal. That's why usability engineers

usually call people *participants* instead of *subjects*; no one likes to feel like a lab rat!

Facilitating a test is a very demanding activity. It takes focus, and it's difficult, if not impossible, for one person to keep the test process running, observe the nuances of the results, and record data. There is no additional energy or time left to greet participants, provide the initial forms, and give them compensation once the testing is complete. Therefore, it is very useful to have additional staff available to handle these functions. Professional testing facilities have support staff.

In some cases, you will need a facility that is geographically separated from your offices. You might decide to do testing in a number of cities intermittently, or you might even need to complete testing in these different cities quite often. In this scenario, it makes sense to have a relationship with a testing facility in each location. These testing facilities are generally set up for marketing studies, but they work well for usability testing. It is also possible to use a conference room in a hotel, but the testing facilities provide such valuable amenities as a greeter, a one-way mirror, built-in sound and video, and usually a more comfortable atmosphere.

Whether you obtain a contract with a professional testing facility or choose to build your own testing space, there are a few advantages associated with obtaining a professional testing facility versus using a simple conference room. Figures 8-1 and 8-2 show the appearance of a typical professional testing facility. Your facility may have a one-way mirror. Most people can tell when you have a one-way mirror, so if your facility has one in place, you should be straightforward about it. With a proper briefing, the mirror works very well. Developers, business owners, and marketing and usability staff can come and observe without disturbing the test. They can discuss what they see and send in their questions to the test facilitator. In place of a one-way mirror you can also use video feeds to adjacent rooms to allow others to observe without disturbing the test.



**Figure 8-1:** *Observer's Side of a Professional Testing Facility Using a One-Way Mirror (see Plate 1)<sup>1</sup>*



**Figure 8-2:** *User's Side of a Professional Testing Facility Using a One-Way Mirror (see Plate 2)<sup>2</sup>*

1. Photo courtesy of The Bureau of Labor Statistics

2. Photo courtesy of The Bureau of Labor Statistics

---

## Recording of Testing Sessions

There is some value to recording the data gathering and testing sessions. Professional facilities have video capability, and the new portable labs allow video as part of their software.

There are two types of tape usage. One common practice is to provide a full videotape of the session for the record. A continuous tape is made of the test, and you end up with many hours of tape. However, if someone says, "I don't believe the user actually did that," you can offer to let him or her see the appropriate portion of the tape. In other cases, a much shorter highlights tape is culled from the full videotaping sessions. This edited video, 5–10 minutes long, shows key findings of the usability testing through the voices and actions of the participants themselves. Carefully selected examples on well-edited highlights tapes often can depoliticize the usability test findings: It is no longer the "opinion" of the tester; it is the voice of the participant. Highlights tapes effectively grip the audience's attention when used as part of the final presentation. This is a very effective practice. There is nothing like showing video of the users in action.

In the past, recording sessions were prohibitively expensive, but with the new "shoebox" usability equipment available today (see Figure 8-3), the cost is much more reasonable. This shoebox equipment includes a TV camera, microphone, monitor, and a remote marker to make it easy to find interesting tape segments. There is in fact no tape, just a high-capacity hard drive to save the data, so it is also far easier to edit and present the results. This ease of use, combined with its reasonable cost, makes the shoebox lab a practical alternative to traditional equipment.

Most labs are moving to digital means of recording to make video editing easier as well. Using this new technology, you can put parts of the video record in the report (see the sample of a test presentation video in Figure 8-4). The lab software lets you record the user's facial expressions and the activity on the screen.

A few labs use a special type of equipment called an eye-tracking device. It lets you track where the user's eye is fixating. You can gain



Figure 8-3: Shoobox Usability Equipment

Highlight 6 of 8 Rec:06/29/02 01:57p - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit

Address [E:\Reports\Cliplibrap\HighL16.htm](http://E:\Reports\Cliplibrap\HighL16.htm)

BuyBailey.com: Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.buybailey.com>

**BUYBAILEY.COM** JUNE 29

801-459-9741 800-628-6104 Fax: 801-459-5100 Fax: 800-628-6994  
 2645 W. 70th St. Suite 200 Salt Lake City, UT 84119-3645, ONYONDEN.COM

The Family Shop | The Reference Store | What's Inside | Opportunities | Find Us Online

Home

- Current Customer
- New Customer
- Hotfiles
- In The Spotlight
- Notes From Sales
- Find Our Plants
- Contact Us
- COLOURPETH PLANTS

**THE BAILEY DIFFERENCE**

**WHAT'S INSIDE**

COMPLETE BAREROOT AND CONTAINER AVAILABILITY

TRADE SHOW SCHEDULE

NOTES FROM SALES AND MARKETING

January 2002

Notes: Has trouble finding the link to availability

Test1: Navigation Issue

Test2: Failure

Test3:

Value1: 3 Value2: 0 Clips: 18

Name: Record:06/29/02 01:57p

Rewind Pause Play

Index Previous Next

Done My Computer

Figure 8-4: Example of a Video Record from a Usability Test (see Plate 3)

a lot of information from this device. You can see users scanning around the page because they are lost or scanning an image because they cannot tell if it is selectable.

Eye-tracking devices are very useful for research purposes. For example, studies have shown that people start scanning in the main area of a Web page and initially ignore the logo, tabs, and left-hand navigation [Schroeder 1998] and that people's eyes are drawn first to areas that have saturated colors (pure bright colors), darker areas, and areas of visual complexity [Najjar 1990].

You do not need an eye-tracking device in order to run an excellent usability test. A good facilitator can see where the user is looking anyway and can supply you with very similar data. An eye-tracking device is expensive and requires setup time, so you probably won't use it for routine usability tests. It may come in handy, however, in a remote usability test since the facilitator will not be physically present with the participant.

---

## Modeling Tools and Software

Most of the important usability work can be completed with a simple office suite. It may help to have a flowcharting package, and you also need software for graphics work, but that's about all the software you need. You also need to be able to use a word processor to document meetings and descriptions, and you need a tool to mock up screens and pages. Which tool you use is not as critical as making sure that the usability staff members are comfortable with the tool and that they do not get distracted or waste time writing "code" to make the screen mock-ups work. Some people prefer a graphics program like Adobe Photoshop, but a presentation tool like Microsoft PowerPoint works just as well. Some usability staff are already facile in a tool such as Microsoft Visio. Whatever tool your staff members already know how to use that allows them to quickly mock up screens and pages is the best tool to use.

Sophisticated modeling tools may or may not be necessary. Available software can assist in the development of very large taskflows

and the modeling of taskflow behavior. An example of this type of software is Micro Saint (a product made by Micro Analysis and Design, Inc.), which supports task modeling. I have seen this software used to good effect in very complex and critical applications, especially in the military design arena. However, I have yet to see this software used to make a difference to commercial Web sites or applications.

Limited modeling tools are available for usability work. You may wish to create your own. At our company, we built an application called the Task Modeler. It is basically a specialized spreadsheet that helps add up the number of clicks, mouse movements, and key-strokes used to complete a task. Using this application on a group of tasks representative of the work to be completed on a given interface provides a good indication of the time it will take an expert user to use the software. This data is important because when measuring the speed of task completion during a usability test, you're measuring only how fast users are during their initial usage, not how fast they would be after extended experience with the interface. During the test, users spend only minutes with the software, so they will not be experts on using the interface. Yet there are many cases when you are designing for expert users. Also, you don't want to make the classic blunder of designing for first usage only. For example, a menu design that can be used easily and quickly by novices is a much better alternative to using commands initially, but you may then find the commands are faster once learned. If the software will be used full time, going with the menu can be a million-dollar mistake. In this case, we built our own tool.

Many companies have purchased tools to track Web sites and provide feedback and statistics on usage. Some of these tools claim to provide usability information and are useful for performing quick checks and validation. For example, there are tools that let you know if your alt text tags are missing (accessibility tools) or if you are using too long a line length. Be wary, however, of tools that track download times for a page, or how many users clicked on a page, or how much time people spent on a Web page, independent of other information. While this information can be useful to know, it can also be misleading. Why did a user spend only 3 seconds on a page? Was it

because (a) the page is poorly designed, (b) the page is well designed and the user got what he or she wanted right away, or (c) the page *before* was poorly designed, so the user clicked on the wrong link? You cannot tell any of this just by reading a report on where people went and how long they stayed. Nothing can replace a trained usability professional evaluating a screen or page or watching and interpreting users performing a task.

---

## Data Gathering and Testing Techniques

Usability data gathering and testing are some of the most valuable tasks your usability team can do. While the phrase “run a usability test” is a general term, keep in mind that there is not a single type of usability test—there are many different types. For example, there are tests of branding, early paper prototyping tests for conceptual design, and later tests on robust working prototypes. You must select the type of test needed for where you are in your development process and then create the correct type of test questionnaires to support the testing.

You can save a lot of money and time by having an initial set of questions and then customizing them as needed for each test. Having a list of standard tests helps to quickly plan the testing, but each test needs its own set of forms, such as video consent forms, facilitator scripts, task instructions, and so on. Defining and creating pre-designed templates can save countless hours. While no template for a given type of test works for all situations, there is certainly value in having a template as part of your infrastructure. Some example template forms include those listed below.

- The **screener** is an essential questionnaire used to select participants for a study. The screener can help eliminate participants who are too sophisticated or too inexperienced. In some cases, a template can be developed and used repeatedly for each study that will access those types of users, though typically the template must be modified for each test.
- **Usability testing routine forms** are a family of forms you need when running usability tests. They are not very exciting, but

they are quite necessary. For example, you must have an informed consent form to get the participant's agreement to participate. Without this form in place, you are in violation of ethics in human research and can be sued. You may also need demographics forms and forms to acknowledge compensation.

As mentioned, there is more than one type of usability test. Below are descriptions of different tests. Which one you use depends on what questions you are trying to answer.

- **Brand perception tests** let you see how the user perceives the current Web site or application. One version of this test is for a single design, and a variant of the test can also be run as a comparison with competitors' designs. Another version of this type of test involves the selection of the best among suggested designs. This test can be conducted with designs from different graphic artists or even different agencies. Regardless of the scenario, the questionnaire for this test must be customized to reflect the company's target brand values. You need to pick the brand values you are interested in testing. What brand values are you looking for, and which do you want to make sure to avoid? Trendy, warm, friendly, sophisticated, "tech-y" . . . you need to customize the questionnaire to get at the data you are interested in.
- If you ask users if they want a given function, they almost always say yes. If you give them a list of potential functions and ask them to rate how important they are, they rate most as very important. But if you give them a list of possible functions and say they can have only three, you get interesting results. This test, called a **functional salience test**, is a great way to identify the relative importance of functions.
- A **test of affordance** determines whether users can tell what they can select on a page. You simply give users a printed copy of a page and tell them, "Circle the items you think you can select and click on." You will see if there are selectable items that users cannot tell are selectable. You will also see if there are items that are not selectable that make users think they *can* select them.
- **"Think aloud" tests** consist of a whole family of tests where the user is told to do a series of tasks, which are observed. Users are

asked to read out loud as they work and tell the facilitator what they are thinking. This is a great way to find problems in a design. You can also estimate how long it will take users to complete tasks.

- The **card sort test** is a useful method if you are trying to find how users categorize the topics in a Web site or application. You create stacks of cards with one item on each card, and then the participants group the cards in a way that makes sense to them. Software can help collect and analyze the groupings used by different participants. The software uses cluster analysis and gives results that can guide the information structure of the design.<sup>3</sup>
- While the card sort test can help guide the design, you can use the **reverse card sort method** to check whether the design worked. You give the participants a list of items and see if they can figure out where to go to find them. If they can find them, the navigational structure is self-evident.
- **Subjective ratings** are a large family of tests that allow users to describe how they feel about your site or application. They decompose or break down the perceptions to allow you to more easily track the cause of problems. For example, you might find that people love the colors but feel that the site is very slow. These findings need to be carefully considered. You might find lots of users saying they want a search facility, but this may actually indicate that there is a problem with the structure of the site. The stated desire for a search facility is often just a symptom of being lost in a poor navigational structure.

---

## The Special Needs of International Testing

International testing is far harder to coordinate than just arranging for facilities and participants in lots of countries and racking up lots of air miles. Test procedures don't work cross-culturally; therefore, international testing takes special capabilities and infrastructure. You need to deal with translation issues and adjust the testing

---

3. IBM's EZSort programs are an example of cluster analysis software. For more information, visit [www3.ibm.com/ibm/easy/eou\\_ext.nsf/Publish/410](http://www3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410).

methodology based on cultural differences. For example, in some cultures it is not polite to criticize, so the usual methods of asking users to think aloud and expecting that they will say what they think is wrong with the product may not work. If you are testing internationally, make sure you leave enough time to deal with these different circumstances.

---

---

### *The Bollywood Method<sup>4</sup>*

*By Apala Lahiri Chavan, Managing Director,  
Human Factors International, Mumbai, India*

The main challenge with usability testing in Asia is that it is impolite to tell someone they have a bad design. It is embarrassing within this culture to admit that you cannot find something, so it is very hard to get feedback.

I conducted a test on a site that offered airline tickets for sale. I used a conventional simulation testing method and got little feedback. I could see that users were not succeeding, but they would not willingly discuss the problems they were experiencing.

I then tried a new method I had developed, called the Bollywood Method. Bollywood is the Hollywood of India and makes far more movies each year than Hollywood does. Bollywood movies are famous for having long and emotionally involved plots. The movies have great pathos and excitement. In applying the Bollywood Method to this testing scenario, I described a dire fantasy situation. I asked each participant to imagine that his or her beautiful, young, and innocent niece is about to be married. But suddenly the family receives news that the prospective groom is a member of the underground. He is a hit man! His whole life story is a sham, *and he is already married!* The participant has sole possession of this evidence and must book airline tickets to Bangalore for him- or herself and the groom's current wife. Time is of the essence!

*(continued)*

---

4. Based on Chavan [2002].

---

---

### *The Bollywood Method (cont.)*

The test participants willingly entered this fantasy, and with great excitement they began the ticket booking process. Even minor difficulties they encountered resulted in immediate and incisive commentary. The participants complained about the button naming and placement. They pointed out the number of extra steps in booking. The fantasy situation gave them license to communicate in a way they never would have under normal evaluation methods.

This method worked well in India and may even be able to be generalized to special situations in North America and other places where participants may hesitate to communicate freely.

---

---

---

## **Recruiting Interview and Testing Participants**

Usability tests typically require fewer participants than marketing research studies because the findings in usability tests are usually qualitative, rather than statistically descriptive. In usability testing, you are not trying to generalize your results and estimate the numbers or percentage of people who feel or would react to a product in a certain way. You are exploring. You are trying to determine whether there are usability issues, what they are if they do exist, and how you might solve them. This means you are trying to delve into the psychology of your users. And this requires that the participants you test are representative of the target population of actual users. You are going to need to find representative users for data gathering and usability testing.

In-house users, while easy to find, aren't usually acceptable participants because they probably care more about the company than the real users do. They see the application as being worthy of additional effort and might exaggerate its value, or they might not flag aspects of the design that make it impractical. They are also familiar with the company's in-house language, concepts, attitude, and mindset,

and they might even have different aesthetic values and perceptions than typical end users.

In one case it is fine to perform tests with in-house users: If you are actually building an application for the internal staff members, it is appropriate to sample them. This is usually a very easy and informal process; the staff members just need to be screened and scheduled.

Lots of market research and usability testing companies have staffs of screeners—clerical-level staff who call lists of potential participants and follow the questions in a special questionnaire (also called a screener, as noted earlier in the chapter). The staff members use the questionnaire to select participants who fit the criteria for the study. Participants are typically offered a fee of \$100 to \$200 each, depending on how stringent the required match criteria are. Some of these facilities have databases of potential participants. This can be convenient, but the lists may be overused. (Some people seem to be making a part-time job out of participating in studies!) You may want a fresher list. To accomplish this, you may need to ask the recruiting firm about the people in their databases. You can shop around for databases and recruiters, and you can specify that the participants must not have been in any studies during the last 12 months. This may make your recruiting more expensive because it may be harder to find participants. If you need general participants, for example, people between the ages of 20 and 60 who purchase goods from the Web at least once every 3 months, it may be relatively easy to find “fresh” participants. If you need people who work in a copy center who have never used a particular type of software, you will pay more for this type of recruiting. It is good to have relationships already set up with companies that can help you recruit participants.

If your user group is current customers, it may be possible to develop a list of customers and have the staff screeners work from that list. This may be easier and more cost-effective than using a recruiting firm. In some cases, you can have internal staff work temporarily as screeners. This costs very little unless you need to hire in-house staff to work as screeners full time. Using in-house screeners saves money over hiring a screener consulting firm, but the in-house screeners will need to be trained. Usability consultants are

already trained and just charge you per project. But having a smooth machine for obtaining study participants helps keep usability work progressing—problems with obtaining participants is the single most common source for the delay of usability projects.

A whole series of deliverable documents result from proper usability work. It is true that some people approach usability without much of a concept of deliverables. They think they can just study the user and good things will happen. That may be true—good things *may* happen. But to make usability work efficient and repeatable requires an organized set of deliverable documents. The deliverables give a clear focus and a set of milestones for usability work. As an example, Table 8-1 lists the major deliverables in The Schaffer Method.

**Table 8-1:** *The Major Deliverables in The Schaffer Method*

<i>Schaffer Method Deliverables</i>	
Expert Review Report	Wireframes
Usability Test Report	Graphic Treatment
Kickoff Presentation	Simulation Test Results
Data Gathering Plan	Functional Specification
Insight of the Data Gathering	Developer Briefing
Contract for Design	Design Modification Log
User Interface Structure	Final Usability Test
Standards Identification	Plan for Process Improvement
Custom Standard	Post Release Evaluation
Module Definition	Localization Assessment
Module Requirements	Interface Translation
Screen Flow, Functions, and Fields	Specification

It takes time to create good deliverables, but they offer several benefits.

- They document that steps in the methodology are actually completed.
- They allow work to be communicated to others, for instance, key stakeholders and development staff.
- They allow work and processes to be repeated.

Most deliverables require several smaller deliverables to create the end product. So, in the end, there are hundreds of deliverables. Imagine that you needed to create these deliverables from scratch for each project, figuring out the appropriate document structure and inventing the style of presentation. The level of investment for this would make usability engineering programs prohibitive in cost and time. If each of the 23 deliverables listed in Table 8-1 took just  $\frac{1}{2}$  a day to create structurally, then you would add  $11\frac{1}{2}$  days to the project.

If usability is to be routine, standard reusable deliverables are indispensable. They help organize the project and save valuable time. Standard deliverables also make it easier for managers to check a project's progress because they know the full set of deliverables to expect. Finally, using standard deliverables also makes it easier to get oriented and to review an unfamiliar project.

---

The value of the tools, templates, and facilities outlined in this chapter is that they save you valuable time. However, it remains critical to pick the items most appropriate for your efforts. It is not sensible to invest in something just because it is a new technology. Refer back to your strategy often, and remember to let your methodology determine your toolkit. The next chapter provides information on another valuable time-saver—the implementation of interface design standards.

