

USABILITY TESTING

A consulting service offered by:

Enterprise Integration Group
2410 San Ramon Valley Blvd., Suite 225
San Ramon, CA 94583
1-888-EIG4IVR
<http://www.eiginc.com>

Introduction and Overview

This document is one of a series of white papers prepared by Enterprise Integration Group (EIG) to help clients understand EIG methods and processes. EIG is a leading provider of design and evaluation products that deliver superior customer self-service solutions to the enterprise.

What is Usability Testing?

Usability testing is an important phase of overall product design. The goal of usability testing is to determine how effectively a given user interface performs when a typical user—with no prior expectations or preconceptions—interacts with it. EIG has found that direct one-on-one usability testing is far more capable at achieving this goal than are focus groups or other market research and product evaluation methods commonly in use.

During testing, users are observed interacting with an application. The test setup is designed to be as faithful as possible to all conditions and user types targeted by the product. Tests are usually videotaped. The sessions are observed—and the tapes subsequently studied—to discover trends and to extract performance data. Elements of the interface that are effective across the test population can be considered validated as

well-designed solutions. Elements that consistently fail are identified, analyzed, and redesigned to improve user interface effectiveness.

EIG can perform usability testing on an existing fielded product, a prototype, or through the use of a wizard simulation.

What is the Value of Usability Testing?

Product quality problems take two general forms: programming bugs and design flaws. It is well known that the costs associated with both problems increases exponentially when the problems are exposed late in the product lifecycle. This means that accurate knowledge of the usability of a given user interface reduces risk in later stages of development. Conversely, discovery of usability problems in late stages of product development lead to costly rewrites or, worse, premature release of faulty products. Either result negatively impacts the cost-reduction or customer-satisfaction goals of IVR and other products.

Usability testing reduces risk and maximizes ROI by injecting realistic and accurate measures into the design and development process. Clients especially value wizard tests because they can occur in the early stages of development, allowing quick appraisal of a proposed design (and complete redesign if warranted). Mid-stage testing of prototypes then calibrate wizard results with higher-fidelity measures, thereby influencing the design while it is still flexible and responsive to change. Late-stage testing—through validation and acceptance methods—provide “just in time” exposure of flaws that are otherwise reported, at great expense, by the customers themselves.

Products that survive rigorous usability testing have a proven track record, and thus demonstrate the predictive validity of the service.

Subject Acquisition

The first and most expensive phase of usability testing is the acquisition and scheduling of a representative sample of users—the so-called “subject pool.” EIG works closely with

the client to ensure that subjects exhibit the demographic characteristics typical of the client's customer base. Clients will occasionally offer their own customers or a list of potential subjects to assist in this process. EIG—with offices in San Ramon, California and Denton, Texas—also has methods for acquiring subjects through school groups, community organizations, and mailing lists.

About Sample Size

A common error in experimental design is the assumption that a large sample size is required for “statistical validity.” Although it is true that small samples can be misleading if experiments are poorly planned or analyzed, the most important attribute of subject selection is “representativeness,” not size. When carefully chosen subjects are truly representative of users that are targeted by the application, a great deal can be learned from even a small number. Conversely, if test subjects bear little demographic relation to the actual users, then even large subject pools deliver poor information.

EIG is competent and professional at subject selection and experimental design.¹ A staff of specialists gather appropriate data, work with clients to iterate test cases, and ensure that the psychology and context of usability tests are valid. The result is superior test results with minimum subject pools. Typical testing sequences use 16 subjects in half-hour sessions, followed by extensive written questionnaires and live interviews. More complex applications or extensive research efforts use twice or occasionally three times as many subjects, longer test sessions, and more sophisticated group-testing techniques.

Methodology

Test Preparation and Facilitation

While subject acquisition is underway, an EIG research scientist must analyze the user interface to understand its intended behavior. This may entail studying the design documents or reverse-engineering a running application. Once the application is

¹ See the EIG publication, “Personnel and Credentials” for more background on EIG capabilities.

understood, the scientist devises a collection of tasks—actions that a typical user might wish the application to perform. These tasks must be described in such a way that the subject can understand the goal easily. Care must be taken, however, that the description of the task not convey the method or user action required to perform the task. Instead, the user must “try things” to accomplish the goal. In this way, the effectiveness of the design under typical uninformed user conditions is measured.

When the time arrives for running the usability tests, subjects arrive at a scheduled time. They are given specific instructions and asked to sign a waiver. They are then exposed to the tasks that they must accomplish and placed in a private room where EIG personnel will not interfere with the subject’s normal and spontaneous behavior. The test room is normally arranged as an office. The subject calls the application and engages in a dialogue using touch tone or speech recognition. The entire interaction is videotaped. Subjects are usually paid. Tests are performed one subject at a time. Subsequent group discussions or debriefings may occur in some testing cases.

Wizard Testing

One form of usability testing, unique to EIG, is live wizard testing. In this process, the “application” under test is a preliminary design as simulated by a live experimenter. The subject calls what he or she thinks is an IVR application. Instead, the telephone (located at another site) is answered by a human experimenter who speaks on behalf of the design.

The experimenter is whimsically called a “wizard” because the caller is expected to “pay no attention to the man behind the curtain.”² The wizard must be specially trained to speak the machine’s part of the dialogue in a consistent and IVR-like manner. The wizard interprets user input—by observing touch tones on a detector or by recognizing caller speech—and uses this input to follow the logic of the design, thereby simulating the proposed user interface.

As a low-fidelity simulation, live wizard testing provides coarse information about call flow and scripting. It is not useful for generating new design ideas, but it is quite effective

² From *The Wizard of Oz* by Frank L. Baum.

at identifying problem areas, revealing logic or sequencing problems, and exposing jargon or unclear wording in prompts.

To increase fidelity, wizard tests may also take advantage of certain software tools common in the human factors arena. Preparation for so-called automated wizard testing deposits pre-recorded prompts into a playback system which is then connected to the telephone. The test itself then proceeds in the same fashion as a live wizard test—the human wizard views touch tones or listens to user speech. To present voice responses to the caller, however, the wizard selects one of a number of audio files and then plays the audio automatically into the telephone. The wizard does not speak. Automated wizard tests exhibit higher fidelity because the pre-recorded prompts are invariant.

Prototype Testing

Once a design is roughly validated with wizard tests, a high-fidelity simulation can be created in the form of a prototype. Clients often engage EIG to create the prototype for the purpose of testing and demonstrating a product prior to the final commitment to develop.³ In addition to development, EIG is often asked to coordinate a formal usability test as part of the larger development effort.

Usability testing of a prototype relies on the same effective subject acquisition methods that EIG uses for other usability tests. Recognizing that the goal of usability tests is to study the behavior of the user interface and not to debug the prototype, test cases are carefully designed to exercise those parts of the application that are thoroughly tested and known to be bug-free. The prototype is also loaded with professionally-recorded prompts for accurate assessment of scripting. Subjects are then acquired, brought to the lab, and asked to call the prototype.

Validation and Acceptance Testing

After an application is complete and ready for Alpha Test, EIG is often asked to perform a final usability test—designed to confirm that the application works as designed and that

³ See the separate white paper, “Prototype Development” for more information.

the design remains effective among real users. This phase of testing is most effective if it follows wizard and prototype testing, as it uncovers details associated with system latency, host-based errors or scripting flaws, or other problems that may not have been replicable in earlier tests.

Occasionally, customers ask EIG to perform usability testing on a new product that has not been previously evaluated. These tests exhibit a considerable degree of risk, and EIG undertakes them with caution. The reason is that usability tests tend to identify usability problems. Unless the organization is prepared to address problems identified by such testing, the effort is wasteful of resource—why run tests if the results will not lead to action? End-game testing is therefore often deferred in favor of direct field experience.

Four-Person Group Tests

One useful form of testing invites four subjects to interact with the application at the same time. The subjects are located in separate private test rooms. Observers view the tests on a split-screen monitor and listen through headsets. The callers are completely independent, requiring four concurrent instances of the application.

When the subjects have completed their individual sessions, they are assembled into a common interview room for an additional group discussion. Bearing slight resemblance to a focus group, the meeting is facilitated by a skilled investigator who interrogates the group using a method known as nominal group technique. The technique compensates for known flaws of focus groups in which the most verbal subject dominates discussion.

Both the individual test sessions and the subsequent group discussion are videotaped.

Deliverables

Physical Deliverables

The deliverables for usability testing are a paper report and a CD containing video clips of relevant events that occurred during testing. Clients also have access to the video library of raw data.

Client Observation Privileges

An important added value of usability testing is participation in the test process itself. Clients come to the lab and sit in a comfortable observation room. Here, in privacy, they may watch subjects on closed circuit TV as they interact with the application. Clients may discuss events as they occur—sharing their observations with EIG specialists who are expert both in the testing methodology and in design methods. Although the end results appear in the final report, test participants often declare that the on-site experience of watching and discussing subject interactions was the most valuable aspect of the usability testing service.

Note that on occasion, observers may invite an especially interesting subject into the observation room for additional interviewing. This, of course, can only occur after the test session is over—subjects are not allowed to see or know about observers before or during the test—but sometimes leads to deep insights regarding what motivated a user to behave a certain way, or what additional thoughts a subject might have regarding the friendliness, efficacy, and overall value of the product.

Timing

Wizard Testing

Wizard testing is the most expensive testing service, as EIG must make use of several highly-specialized resources for the duration of the planning and testing effort. Preliminary preparation occupies several days. Testing itself generally requires two or

three long days. Post-test analysis and video clip extraction is then followed by a final report, requiring some two weeks or more of additional effort.

Typical Calendar Time: Three to Six Weeks

Prototype Testing, Validation and Acceptance Testing

Usability testing of an application that runs automatically—whether a prototype or a finished application—is less expensive than wizard testing. Other than the additional work of the wizard, however, the effort is roughly the same. Preliminary preparation occupies several days. Testing itself generally requires two or three long days. Post-test analysis and video clip extraction is then followed by a final report, requiring some two weeks or more of additional effort.

Typical Calendar Time: Three to Six Weeks