# Moderated Usability Test

Usability testing is a method in which representative users are recruited to participate in an evaluation of a system’s user interface to assess the degree to which the system meets usability objectives. These objectives typically relate to the system’s usefulness, efficiency, effectiveness, and user satisfaction.

Usability testing is performed iteratively throughout the design and development process. Key goals of testing are to: inform design decisions with targeted user data; estimate user performance within a particular system design (e.g., the amount of time spent trying to complete a particular task, task success rates, and the frequency of use errors); identify usability problems and frustrations that may make the system less efficient, effective, useful and satisfying for end users; and help ensure that the system supports the safe delivery of patient care, prior to release. An evaluation involves developing a test plan, recruiting representative users to participate, conducting test sessions, analyzing results, and disseminating findings. A moderated usability testing session should be conducted by a trained (usability test) facilitator, or moderator, and may occur in a laboratory setting, in the field, in a simulated environment, or remotely/virtually. Depending on the project team’s purpose for testing and when testing is conducted, the evaluation goals, performance measures, and methods may vary.

Usability testing is commonly described as either formative or summative in style. Formative evaluations are conducted early in the design process to help "form" the design for a product or service by detecting and eliminating usability problems. The project team can use these findings to immediately improve the design of the product or service and refine the development specifications. Formative evaluation results can be less formal than in summative evaluations, as suits the needs of designers, developers, project managers, and other project stakeholders. Summative evaluations are generally more formal and are conducted later in the design process, when a system is substantially developed, to verify and validate that the design requirements and corresponding usability criteria have been met prior to deployment. Following good engineering practices and including formative testing early and often increases the likelihood that the summative testing will meet the desired usability levels.

An essential distinguishing factor of usability testing, compared to other usability inspection methods, is the inclusion of representative end users attempting realistic tasks with the system or a representation of the system. As in all user-focused methods, it is important to recruit participants that are either actual end-users or representative of and very similar to the expected end-users of the system. The wrong participants can lead to unreliable or misleading results that can put projects at risk. HFE guidance on [participant recruiting and use](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Working%20with%20Study%20Participants.aspx) can help avoid problems in this area.

Although some sources lump usability inspection methods into the “formative” evaluation category, using the term to describe many methods that are suitable early in the design process, HFE primarily uses the “formative” moniker in relation to testing with users and representative tasks, as previously described. Several types of usability tests used by HFE include:

Formative Exploratory Test – Exploratory testing is a user-based evaluation of unimplemented technology or software concepts conducted early in the design process. The technique is used to uncover interaction design issues during an exploratory phase of a project; it often involves having the project team simulate some or all of the responses of the system. This approach can be used to explore design and usability of emerging technologies such as natural language applications or pervasive computing applications. It is also well suited to testing paper or other low fidelity prototypes early in the design process. The facilitator frequently asks participants probing questions or encourages participants to “[think aloud​](https://vaww.portal2.va.gov/sites/humanfactors/BoKContent/Conducting%20Test%20Sessions%20-%20Think%20Aloud.docx)” while they interact with the system to elicit rich qualitative data that will help inform design decisions.

**Formative Assessment Test** – An assessment test compares the usability of a design against specific measureable performance and/or satisfaction goals, much like a summative evaluation, but with a lower level of effort and in an earlier development phase than a summative evaluation. The facilitator may employ the “think aloud” technique, which is frequently used with exploratory testing, depending on the goals of the specific evaluation and phase of development.

**A/B Test** - A comparison (contrast) test that is carried out by having a group of targeted users perform representative tasks on two or more alternative system designs or prototypes for the purpose of comparing user performance between the designs. Using within-group (i.e., comparing the same users’ performance across all interfaces for a particular system) or between-group comparisons (i.e., each participant is only exposed to one interface, and their performance is compared to other participants, who are the same type of targeted user, using another interface) will provide project teams with empirical data for design decisions. Comparative testing methods can be incorporated into the different usability test methods, which are described above, and applied at multiple phases of the design process. Usability.gov and others use the “A/B test” moniker for unmoderated tests done on a live site, but HFE uses the more traditional meaning described here.

**Summative Test** - Summative usability evaluation is a form of validation that is performed late in the design process, when the system or application is significantly developed and attention turns toward usability and performance metrics, for example in field environments where the application or system is intended to be used. The goal of this type of an evaluation is to determine how well the developed product meets defined usability or performance objectives and establish usability benchmarks for the system for future enhancements or redesign. Summative usability tests generally focus more on quantitative measures, such as task success, time on task, use errors, and user satisfaction scores, although qualitative data is often collected during the debrief interview following completion of the tasks. Tests conducted as part of summative usability testing require that representatives from each group of targeted users, who will be using the system, carry out the predetermined tasks independently by using the intended combination of hardware and software. For example, if the application is to be used by doctors and nurses, both groups would have to be included in the evaluation. Each distinct user population carries out the tasks using the intended combination of hardware and software, either within their current workflow or in an environment in which the workflow is simulated. Testing must allow for the time and coordination for the system to be used in multiple scenarios, which are commonly encountered or critical to the successful and safe performance of the targeted users’ tasks for which the system is designed.

## Moderated Usability Testing Benefits

* Identifies major usability problems that are frequently missed by inspection methods.
* Elicits qualitative feedback, particularly during early exploratory testing, that allows the project team to effectively evaluate how well a proposed design solution meets users’ needs.
* Uncovers usability problems that are driven by user-specific expectations, needs, and abilities that may be difficult to discover without an experienced facilitator’s well-timed probing questions.
* Evaluates how well the system fits into the targeted users’ work environment (e.g., impact on workflows, workload, and efficiency) and, consequently, provides feedback that reveals potential design flaws and other issues (e.g., workflow changes resulting in inefficiencies, disruptions, increased workload, and workarounds).
* Leverages experienced test facilitators’ ability to elicit feedback interactively from users, probe to understand why they had problems, and draw out potential design ideas from participants.
* Enables business office representatives and project team members (e.g., UI Designers and developers) to easily observe the evaluation sessions or access recordings.
* Allows for cost-effective testing of low and medium-fidelity prototypes.
* Produces video clips, screen shots, and quotes from test sessions to illustrate potential issues and concerns.
* Obtains reliable measures of users’ performance (e.g., actions and inactions; tasks and goals effectively performed/completed/achieved), efficiency and satisfaction.
* Provides statistical measures (summative testing) of usability and performance (e.g., success rate, average time to complete task, etc.). The desired levels of usability are established at the beginning of the development lifecycle, providing clear criteria for the evaluation.

## Moderated Usability Testing Limitations

* Requires a solid foundation of quality user research methods to understand the context of use, realistic use scenarios, and user needs early in the design process. Usability testing alone is insufficient to inform an efficient, effective, satisfying and safe user experience since there are inherent limitations in simulation or laboratory testing that may miss key factors present in real world use environments.
* Runs the risk of focusing too narrowly on a system’s ease of use without properly evaluating the system’s usefulness to end users. A well-designed evaluation should also assess the utility of a system since this is a critical factor for successful user adoption (i.e., users are unlikely to use a system, regardless of how easy it is to use, if the tools are not appropriate or advantageous in the user’s actual work environment).
* Introduces and involves an inherently artificial situation; demonstrated high, or at least acceptable, levels of usability (e.g., high levels of users’ task success rates, efficiency, and satisfaction) in a laboratory or simulated environment does not guarantee usability in actual work environments and conditions (Rubin & Chisnell, 2008).
* Requires potentially time-consuming and resource intensive efforts to be thorough, valid and reliable (e.g., evaluate all of the key tasks, goals and scenarios for which the system is designed).
* Requires representation of all targeted user types in the evaluation, particularly in summative testing; it may be difficult to recruit a sufficient sample of each user type, whose availability may be quite limited, requiring a substantial amount of compensation or another worker to relieve him or her during the evaluation session.
* Contains the risk that not all problems will be found with small samples of users (Faulkner, 2003).
* Requires potentially complex technical setup, access to domain experts and additional time for troubleshooting and debugging.
* Requires a basic level of test facilitation experience to avoid influencing the results by using too many hints, asking biased questions, providing nonverbal cues about the tasks, or interrupting the users at inappropriate times (and, consequently, disrupting users’ task processes, thoughts, actions and decisions).
* Results in potentially higher costs or substantial delays if a significant usability problem is discovered late in the development lifecycle, such as during summative testing. Since the system is all but complete at this point, redesign and redevelopment to address the issue may be infeasible or otherwise unacceptable. Conducting early and iterative formative testing and using sound HFE methodology significantly mitigates this risk.

​

## Study Execution

|  |  |
| --- | --- |
| **Milestone** | **​Owner** |
| ​Initiate Kick-off Call | HFE |
| Develop Proposal | ​                                  HFE |
| Finalize objectives and performance measures | HFE, ​Business Office |
| ​Develop Tasks/scenarios | ​Business Office |
| ​Verify that tasks meet objectives | ​                                  HFE |
| Provide system access or a representation of the system | ​                         Business Office​ |
| ​Create test data | ​Business Office |
| ​Verify Environment | ​HFE |
| ​Recruit Participants | ​Business Office​ |
| ​Schedule Participants | HFE |
| ​Pilot test the study materials and procedures | ​HFE |
| ​Conduct usability assessment | HFE |
| Attend results briefing/ collaborate on solutions | HFE​, Business Office, Developer |
| ​Analyze findings, write report | HFE |
| ​[Conduct After Action Review](https://vaww.portal2.va.gov/sites/humanfactors/SitePages/AAR%20Process.aspx) | HFE |

## Outcomes

* A results briefing to discuss what solutions are possible for usability issues, within the given constraints.
* Summary report of the results of the satisfaction questionnaire, task time and effectiveness measures (e.g., accuracy, completeness, and efficiency), if used.
* A list of identified usability issues and concerns, categorized by severity and frequency.

​

The type of objective and subjective data collected during testing may include:

## User Effectiveness

* Completion rate – the percentage of participants who completely and correctly achieve each task goal.
* Errors – instances where tasks were not completed successfully; types and number of use errors; number of repeated use errors.
* Assists – when direct procedural help is provided by the facilitator.
* Number of design issues that lead to user confusion.

## ​User Efficiency

* Task time – mean times to complete tasks with the range and standard deviation of times across participants.
* Mean time on task – the percentage of users who were successful for each unit of time.
* Sequence and number of steps to complete a task.

## User Satisfaction

* Analysis of standardized measure of user satisfaction with the system.
* Ratings of ease of performing a task.

Users’ qualitative comments.

|  |  |
| --- | --- |
| **Phase of Development​** | |
| ​ | ​Planning, Scoping & Definition |
| ​ | ​Requirements Gathering |
| **✔** | **​Early Design** |
| **​✔** | **​Detailed Design & Development** |
| **​✔** | **​Field Testing** |
| ​ | ​Deployment |
| **✔** | **​Post-Deployment** |

|  |  |
| --- | --- |
| **​Study Characteristics ​** | |
| ​**Timeframe** | ​5-6 weeks |
| **​Level of Effort** | High |
| ​**Data Collection** | Behavioral, Attitudinal |
| ​**Data Reporting** | ​Quantitative, Qualitative |

|  |  |
| --- | --- |
| **​Related Methods ​** | |
| **Derived from** | ​-- |
| **​Complimentary Methods** | [**Heuristic Evaluation**](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Heuristic%20Evaluation.aspx), [**Cognitive Walkthrough**](https://vaww.portal2.va.gov/sites/humanfactors/BoKSitePages/Methods/Cognitive%20Walkthrough.aspx) |
| **​Similar Methods** | [**User Experience Assessment**](https://vaww.portal2.va.gov/sites/humanfactors/BoKSitePages/Methods/User%20Experience%20Assessment.aspx) |
| **​Follow-Up** | [**--**](https://vaww.portal2.va.gov/sites/humanfactors/HFBoK/SitePages/Summative%20Usability%20Test.aspx) |

|  |
| --- |
| **​Referenced on this page...** |
| Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers, 35(3), 379-383.  Rubin, J., & Chisnell, D. (2008). Handbook of Usability Testing. Indianapolis, IN: Wiley Publishing, Inc. |