Usability evaluation

# Definition

Usability evaluation consists in testing a product or a service with the representative users under realistic conditions. The purpose of this method is to detect the usability problems by collecting qualitative and quantitative data such as success rate, task time and user satisfaction with the product in order solve them and improve the system. The typical procedure of the evaluation is that the users are supposed to accomplish some representative tasks with the system while the tester watch, listen and takes notes about the performance and the difficulties the users encounter.

There are different techniques to perform usability evaluation:

* simple observation
* constructive observation
* think-aloud evaluation

5 users qualitative

more quantitative

# Benefits

The main and most valuable benefit of performing usability test is that it detects difficulties and problems before they are implemented:

🡪 the earlier they are detected and solved, the less expensive, in terms of time and money, the cost of the system development is

# Best practices

* Test early test often

In order to perform an effective usability evaluation, it is crucial to prepare a good plan, to recruit the participants and finally analyse and report what has been found.

### Planning

* It is important that the users, tasks and environment used for the test are representative of the intended context of use.
* Select the most important tasks and user groups to be tested (e.g. the most frequent or the most critical).
* Select users who are representative of each user group. 3-5 users are sufficient to identify problems. 8 or more users of each type are required for reliable measures.
* Produce a task [scenario](http://www.usabilitynet.org/tools/scenarios.htm) and input data and write instructions for the user (tell the user what to achieve, not how to do it).
* Plan sessions allowing time for giving instructions, running the test, answering a questionnaire, and a post-test interview.
* Invite developers to observe the sessions if possible. An alternative is to videotape the sessions, and show developers edited clips of the usability problems.
* Two administrators are normally required to share the activities of instructing and interviewing the user, operating video equipment (if used), noting problems, and speaking to any observers.
* If possible use one room for testing, linked by video to another room for observation.
* If usability measures are required, observe the user without making any comments.
* If measures are not required, prompt the user to explain their interpretation of the contents of each screen and their reason for making choices.
* Test the evaluation before running it (for example with your internal staff). That allows you to:
  + Test the equipment
  + Provides practice for the facilitator and note-takers
  + Get a good sense whether your questions and scenarios are clear to the participant
  + Make any last minute adjustments

## Recruiting Usability Test Participants

It is important to recruit users that are representative of the user group(s) targeted for the system. You can ask to your internal staff to be participants only if they were not involved during design and development sessions of the system and if they represent members of the target group(s). Remember that you don’t random participants but the right target user.

### How many participants are enough?

Nielsen outlines the [number of participants](http://www.nngroup.com/articles/how-many-test-users/) [xternal Link Disclosure ](http://www.hhs.gov/disclaimer.html) that you need based on a number of case studies:

* **Usability Tests**: test **5 users** lets you find almost as many usability problems as you'd find using many more test participants.
* **Quantitative studies** (aiming at statistics, not insights): test at least **20 users** to get statistically significant numbers; tight confidence intervals require more users.

## Costs of Recruitment

Recruiters generally charge a fee for each participant “successfully” recruited. A successful recruit is one that meets the criteria, appears for testing and is able to complete the test. A good recruiter will screen, schedule and remind the participants about their test appointment to assure all of their recruits are successful.

If needs be, you may also engage the recruiter to handle additional administrative duties such as  administering incentives for participants (i.e. gifts or money), and in some cases travel/parking expenses. There will be a fee for additional services so it will be best to discuss any additional services needed by your team during your initial discussions with the recruiter.

## Working with a Recruiter

If the team has access to representative users you can recruit from those individuals.  If the team does not have access to representative users, you will have to hire a commercial recruiting company. Most recruiting companies require two to three weeks to find and schedule the necessary number and types of participants. Here is the basic information you should have available when you speak to a recruiter or request a quote for their services:

* How many participants you will need; total and within specified groups.
* The location, date and time for testing. It is helpful to provide a draft of your screeners (government-focused template and non-government template), a detailed schedule and a map or directions to your facility, if possible.
* How long each session will take.
* If you will be compensating for testing, how much [and in what format; cash, check or gift cards or certificates), and what if anything else you will be providing (i.e. travel or parking).

**An example communiqué with a recruiter:**

Details:

* Date and Time: Friday, July 19th and Monday, July 22nd – Friday, July 26th. (The schedule has the specific time periods on those days we are looking to fill.)
* Location: On-site: (Map attached)
* Length of Each Session: 1-hour usability test sessions
* Compensation for Participants: $\_\_\_ for the session (We will not provide additional compensation for travel or parking.)

As to the participants - Total (18):

* Non-smokers/Non-tobacco users with smokers in the house (5 participants)
* Non-smokers/Non-tobacco users without smokers in the house (4 participants)
* Current smoker or tobacco user (4 participants)
* Former smoker or tobacco user (5 participants)
* A mix of genders
* A mix of ethnicities
* A mix of educational background
* I would like the following ages to be represented in each audience segment:
  + 18-30
  + 31-50
  + 50+

## Compensating a Participant

When determining how and how much to compensate participants for their time it is important to determine what methods and amounts have been used previously.  Your recruiter may be helpful in learning what is typical for the tests on which they have recruited. Your colleagues may also be helpful, if they have either conducted or participated in testing.

Some things to consider when compensating participants:

* If your participants are **federal employees**, you cannot pay them for their time.
* If your participants are **non-federal employees**, the mode of compensation should be in-demand by potential participants.  Money in any form is generally acceptable but it may be more convenient to provide gift cards or certificates for online or local vendors. Keep in mind that purchases from an online vendor will generally charge your participants for shipping.  You may want to adjust the compensation accordingly.
* If you are doing **remote testing**, you may want to consider an electronic mode of compensation such as an eCertificate to an online vendor.
* Remember to **provide a receipt** for your participants (one for [adults](http://www.usability.gov/how-to-and-tools/resources/templates/receipt-form-usability-test-compensation-adult.html) and one for [minors](http://www.usability.gov/how-to-and-tools/resources/templates/receipt-form-usability-test-compensation-minor.html) are included in the templates section) to sign for the purposes of:
  1. Showing that they received the compensation and
  2. Providing documentation to your accounting department or personnel.

### Running sessions

* Welcome the user, and give the task instructions.
* Observe the interaction and note any problems encountered.
* Time each task.
* Interview the user to gain general opinions and to ask about specific problems encountered.
* Assess the results of the task for accuracy and completeness.

## Best Practices

* Treat participants with respect and make them feel comfortable.
* Remember that you are testing the site not the users. Help them understand that they are helping us test the prototype or the actual system.
* Remain neutral – you are there to listen and watch. If the participant asks a question, reply with “What do you think?” or “I am interested in what you would do.”
* Do not jump in and help participants immediately and do not lead the participant. If the participant gives up and asks for help, you must decide whether to end the scenario, give a hint, or give more substantial help.
* The team should decide how much of a hint you will give and how long you will allow the participants to work on a scenario when they are clearly going down an unproductive path.
* Take good notes. Note-takers should capture what the participant did in as much detail as possible as well as what they say (in their words).  The better the notes are that are taken during the session, the easier the analysis will be.
* Measure both performance and subjective (preference) metrics. People's performance and preference do not always match. Often users will perform poorly but their subjective ratings are very high. Conversely, they may perform well but subjective ratings are very low.
  + Performance measures include: success, time, errors, etc.
  + Subjective measures include: user's self reported satisfaction and comfort ratings.

## Running a Usability Test

Once you have [planned your test](http://www.usability.gov/how-to-and-tools/methods/planning-usability-testing.html) and [recruited your test participants](http://www.usability.gov/how-to-and-tools/methods/recruiting-usability-test-participants.html), it’s time to get ready to conduct your test.  To do so, you’ll want to think about which moderating technique is right for your test, set up your space and equipment, and make sure that you do a pilot test prior to testing with actual participants.

Choosing a Moderating Technique

In her [Moderating Usability Tests article](http://www.usability.gov/get-involved/blog/2013/04/moderating-usability-tests.html), Jen Romano Bergstrom notes that choosing the best moderating technique for your test depends on your session goals.  Some common moderating techniques include:

* **Concurrent Think Aloud (CTA)** is used to understand participants’ thoughts as they interact with a product by having them think aloud while they work. The goal is to encourage participants to keep a running stream of consciousness as they work.
* In **Retrospective Think Aloud (RTA)**, the moderator asks participants to retrace their steps when the session is complete. Often participants watch a video replay of their actions, which may or may not contain eye-gaze patterns.
* **Concurrent Probing (CP)** requires that as participants work on tasks—when they say something interesting or do something unique, the researcher asks follow-up questions.
* **Retrospective Probing (RP)** requires waiting until the session is complete and then asking questions about the participant’s thoughts and actions. Researchers often use RP in conjunction with other methods—as the participant makes comments or actions, the researcher takes notes and follows up with additional questions at the end of the session.

It’s important to weigh the pros and cons when you are trying to decide which technique to use:

|  |  |  |
| --- | --- | --- |
| **Techniques** | **Pros** | **Cons** |
| **Concurrent Think Aloud (CTA)** | * Understand participants’ thoughts as they occur and as they attempt to work through issues they encounter * Elicit real-time feedback and emotional responses | * Can interfere with usability metrics, such as accuracy and time on task |
| **Retrospective Think Aloud (RTA)** | * Does not interfere with usability metrics | * Overall session length increases * Difficulty in remembering thoughts from up to an hour before = poor data |
| **Concurrent Probing (CP)** | * Understand participants’ thoughts as they attempt to work through a task | * Interferes with natural thought process and progression that participants would make on their own, if uninterrupted |
| **Retrospective Probing (RP)** | * Does not interfere with usability metrics | * Difficulty in remembering = poor data |

## Example Usability Test Session

Here is an example test session.

1. The facilitator will welcome the participant and explain the test session, ask the participant to sign the release form, and ask any pre-test or demographic questions.
2. The facilitator explains thinking aloud and asks if the participant has any additional questions. The facilitator explains where to start.
3. The participant reads the task scenario aloud and begins working on the scenario while they think aloud.
4. The note-takers take notes of the participant’s behaviours, comments, errors and completion (success or failure) on each task.
5. The session continues until all task scenarios are completed or time allotted has elapsed.
6. The facilitator either asks the end-of session subjective questions or sends them to an online survey, thanks the participant, gives the participant the agreed-on incentive, and escorts them from the testing environment.

The facilitator then resets the materials and equipment, speaks briefly with the observers and waits for the next participant to arrive.

## Output

* Produce a list of usability problems, categorised by importance (use post-it-notes to [sort](http://www.usabilitynet.org/tools/affinity.htm) the problems), and an overview of the types of problems encountered.
* Arrange a meeting with the project manager and developer to discuss whether and how each problem can be fixed.
* If measures have been taken, summarise the results of the satisfaction questionnaire, task time and effectiveness (accuracy and completeness) measures.

## Reporting Usability Test Results

When reporting results from a usability test, you should focus primarily on your findings and recommendations that are differentiated by levels of severity.  Include the pertinent information from the test plan and present just enough detail so that the method is identifiable.  Keep the sections short, use tables to display the metrics, and use visual examples to demonstrate problem areas, when possible.

## Data Analyses

At the end of usability testing you will have collected several types of data depending on the [metrics you identified in your test plan](http://www.usability.gov/how-to-and-tools/methods/planning-usability-testing.html). When analyzing the data you’ve collected, read through the notes carefully looking for patterns and be sure to add a description of each of the problems. Looks for trends and keep a count of problems that occurred across participants.

|  |  |
| --- | --- |
| **Quantitative Data** | **Qualitative Data** |
| * Enter the data in a spreadsheet to record data or make calculations such as:   + **Success rates**   + **Task time**   + **Error rates**   + **Satisfaction questionnaire ratings** * You may want to add participant’s demographic data so that you can sort by demographics to see if any of the data differ by the demographic variables. * Make sure you identify the task scenarios for each of the metrics. | * Record data related to:   + **Observations about pathways participants took**   + **Problems experienced**   + **Comments/recommendations**   + **Answers to open-ended questions** * Make sure your problem statements are exact and concise. For example:   + Good problem statement: Clicked on link to Research instead of Clinical Trials.   + Poor problem statement: Clicked on wrong link.   + Poor problem statement: Was confused about links. |

## Reporting Severity Levels of Problems

As you are reviewing the data, consider how global the problem is throughout the site and how severe (or serious) the problem is.  Your findings may have implications for other pages in the site (global). For example, you may find that participants could not find what they needed on the page because of text density. You could say that just that page needed to be fixed but you should also consider how many other pages are equally dense with text.

Some problems contribute more to participants not being able to complete the scenarios than others. To help differentiate, you should note the severity of the problems on a three- or four-point scale. For example:

* **Critical**:  If we do not fix this, users will not be able to complete the scenario.
* **Serious**:  Many users will be frustrated if we do not fix this; they may give up.
* **Minor**:  Users are annoyed, but this does not keep them from completing the scenario. This should be revisited later.

## Writing the Usability Test Report

In general, your report should include a background summary, your methodology, test results, findings and recommendations.  There are a number of [report templates](http://www.usability.gov/how-to-and-tools/resources/templates.html) that you may adapt to assist you in reporting your findings.

* **Background Summary**:  Include a brief summary including what you tested (website or web application), where and when the test was held, equipment information, what you did during the test (include all testing materials as an appendix), the testing team, and a brief description of the problems encountered as well as what worked well.
* **Methodology**:  Include the test methodology so that others can recreate the test. Explain how you conducted the test by describing the test sessions, the type of interface tested, metrics collected, and an overview of task scenarios.  Describe the participants and provide summary tables of the background/demographic questionnaire responses (e.g., age, professions, internet usage, site visited, etc.). Provide brief summaries of the demographic data, but do not include the full names of the participants
* **Test Results**:  Include an analysis of what the facilitator and data loggers recorded. Describe the tasks that had the highest and lowest completion rates.  Provide a summary of the successful task completion rates by participant, task, and average success rate by task and show the data in a table. Follow the same model for all metrics.  Depending on the metrics you collected you may want to show the:
  + Number and percent of participants who completed each scenario, and all scenarios (a bar chart often works well for this)
  + Average time taken to complete each scenario for those who completed the scenario
  + Satisfaction results
  + Participant comments can be included if they are illustrative.
* **Findings and Recommendations**:  List your findings and recommendations using all your data (quantitative and qualitative, notes and spreadsheets). Each finding should have a basis in data—in what you actually saw and heard.  You may want to have just one overall list of findings and recommendations or you may want to have findings and recommendations scenario by scenario, or you may want to have both a list of major findings and recommendations that cut across scenarios as well as a scenario-by-scenario report.  Keep in mind:
  + Although most usability test reports focus on problems, it is also useful to report positive findings. What is working well must be maintained through further development.
  + An entirely negative report can be disheartening; it helps the team to know when there is a lot about the Web site that is going well.
  + Each finding should include as specific a statement of the situation as possible.
  + Each finding (or group of related findings) should include recommendations on what to do.

## Incorporating Visuals to Illustrate Specific Points

You can make the report both more informative and more interesting by including visual content. You may consider including:

* **Screen shots to readers visualize what you were testing.** Include parts of screens to illustrate specific areas that are working particularly well or that are causing problems for users.
* **Short video clips to illustrate specific points**, if you are presenting the report electronically and the readers of the report have the technology available to see video clips. People who did not observe the actual test sessions are often most convinced of problems and the need to fix them by watching and listening to relevant video clips.

## Implement and Retest

For a usability test to have any value, you must use what you learn to improve the site. You may not be able to implement all the recommendations. Developing any product is a series of trade-offs in which you balance schedule, budget, people's availability, and the changes that are needed.  If you cannot implement all the recommendations, develop priorities based on fixing the most global and serious problems. As you prioritize, push to get the changes that users need.

Remember that the cost of supporting users of a poorly-designed site is much greater than the cost of fixing the site while it is still being developed.

## You Do Not Need a Formal Lab

Effective Usability Testing does not require a formal usability lab for testing. You can do effective usability testing in any of these settings:

* Fixed laboratory having two or three connected rooms outfitted with audio-visual equipment
* Room with portable recording equipment
* Room with no recording equipment, as long as someone is observing the user and taking notes
* Remotely, with the user in a different location (either moderated or unmoderated)

It’s important to keep in mind that usability testing is not just a milestone to be checked off on the project schedule. The team should have a goal for why they are testing and then implement the results.

t's important to test users individually and let them solve any problems on their own. If you help

them or direct their attention to any particular part of the screen, you have contaminated the test

results.

To identify a design's most important usabilit

y problems,

testing 5 users

is typically enough.

Rather than run a big, expensive study, it's a better use of resources to run man

y small tests

and revise the design between each one so you can fix the usability flaws as you identify

them.

Iterative design

is the best way to increase the quality of user experience. The more

versions and interface ideas you test with users, the better.

User testing is different from

focus groups

, which are a poor way of evaluating design

usability. Focus groups have a place in market research, but to evaluate interaction designs

you must closely observe individual users as they perform tasks wit

h the user interface.

Listening to what people say

is misleading: you have to watch what they actually do.

* Decide on the tasks
* Formulate the tasks
* Be tactful in presenting the order of the tasks

Dumas and Fox (2008, p1131) provide a very good summary of the kind of tasks that are likely to be involved in usability testing. It is in line with those that we used in our testing sessions in most contexts. These include:

* tasks that are important, such as frequently performed tasks or tasks that relate to important functions;
* tasks where evaluators predict users will have difficulties;
* tasks that enable a more thorough examination of the system, such as those that can only be accomplished by navigating to the bottom of the system hierarchy, or tasks that have multi-links or shortcuts;
* tasks that influence business goals;
* tasks that examine the re-designed areas;
* tasks that relate to newly-added features.

http://design.canonical.com/2013/08/usability-testing-how-do-we-design-effective-tasks/

# Examples

<http://www.usabilitynet.org/tools/testing.htm>

<http://www.usability.gov/what-and-why/glossary/usability-evaluation-or-assessment.html>

<http://www.usability.gov/how-to-and-tools/methods/usability-testing.html>

<http://www.usability.gov/how-to-and-tools/methods/reporting-usability-test-results.html>

<http://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html>

<http://www.usability.gov/how-to-and-tools/methods/recruiting-usability-test-participants.html>

http://www.nngroup.com/articles/how-many-test-users/

libro 424