These notes are from usability.gov.
They were gathered by Anthony Hornof in March, 2023.
The content is 100% up-to-date with regards to usability testing.
To keep the document short, I removed a couple pages.
The URLs for each web page are indicated on each page.

The sections include:

Usability Testing

Usability testing refers to evaluating a product or service by testing it with representative users. Typically, during a test, participants will try to complete typical tasks while observers watch, listen and takes notes. The goal is to identify any usability problems, collect qualitative and quantitative data and determine the participant's satisfaction with the product.

To run an effective usability test, you need to develop a solid test plan, recruit participants, and then analyze and report your findings.

Benefits of Usability Testing

Usability testing lets the design and development teams identify problems before they are coded. The earlier issues are identified and fixed, the less expensive the fixes will be in terms of both staff time and possible impact to the schedule. During a usability test, you will:

- Learn if participants are able to complete specified tasks successfully and

- Identify how long it takes to complete specified tasks

- Find out how satisfied participants are with your Web site or other product

- Identify changes required to improve user performance and satisfaction

- And analyze the performance to see if it meets your usability objectives

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)

Factors Affecting Cost

Your testing costs depend on

- Type of testing performed

- Size of the team assembled for testing

- Number of participants for testing

- Number of days you will be testing

Remember to budget for more than one usability test. Building usability into a Web site (or any product) is an iterative process. Consider these elements when budgeting for usability testing:

- **Time**: You will need time to plan the usability test. It will take the usability specialist and the team time to become familiar with the site and pilot test the test scenarios. Be sure to budget in time for this test prep as well as running tests, analyzing the data, writing the report, and presenting the findings.

- **Recruiting Costs**: Consider how or where you will recruit your participants. You will either need to allow for staff time to recruit or engage a recruiting firm to schedule participants for you based on the requirements.

- **Participant Compensation**: If you will be compensating participants for their time or travel, factor that into your testing budget.

- **Rental Costs**: If you do not have monitoring or recording equipment, you will need to budget for rental costs for the lab or other equipment. You may also need to secure a location for testing, a conference room for example, so consider this as well.

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.

Planning a Usability Test

One of the first steps in each round of usability testing is to develop a plan for the test. The purpose of the plan is to document what you are going to do, how you are going to conduct the test, what metrics you are going to capture, number of participants you are going to test, and what scenarios you will use.

Typically, the usability specialist meets with the site or product owner and members of the development team to decide on the major elements of the plan. Often, the usability specialist then drafts the plan, which circulates to management and the rest of the team. Once everyone has commented and a final plan agreed upon, the usability specialist revises the written plan to reflect the final decisions.

Elements of a Test Plan
You will need to include these elements in the usability test plan.

- **Scope**: Indicate what you are testing: Give the name of the Web site, Web application, or other product. Specify how much of the product the test will cover (e.g. the prototype as of a specific date; the navigation; navigation and content).

- **Purpose**: Identify the concerns, questions, and goals for this test. These can be quite broad; for example, "Can users navigate to important information from the prototype's home page?" They can be quite specific; for example, "Will users easily find the search box in its present location?" In each round of testing, you will probably have several general and several specific concerns to focus on. Your concerns should drive the scenarios you choose for the usability test.

- **Schedule & Location**: Indicate when and where you will do the test. If you have the schedule set, you may want to be specific about how many sessions you will hold in a day and exactly what times the sessions will be.

- **Sessions**: You will want to describe the sessions, the length of the sessions (typically one hour to 90 minutes). When scheduling participants, remember to leave time, usually 30 minutes, between session to reset the environment, to briefly review the session with observer(s) and to allow a cushion for sessions that might end a little late or participants who might arrive a little late

- **Equipment**: Indicate the type of equipment you will be using in the test; desktop, laptop, mobile/Smartphone. If pertinent, include information about the monitor size and resolution, operating system, browser etc. Also indicate if you are planning on recording or audio taping the test sessions

or using any special usability testing and/or accessibility tools.

- **Participants**: Indicate the number and types of participants to be tested you will be recruiting. Describe how these participants were or will be recruited and consider including the screener as part of the appendix.

- **Scenarios**: Indicate the number and types of tasks included in testing. Typically, for a 60 min. test, you should end up with approximately 10 (+/-2) scenarios for desktop or laptop testing and 8 (+/- 2) scenarios for a mobile/smartphone test. You may want to include more in the test plan so the team can choose the appropriate tasks.

- **Metrics**: Subjective metrics: Include the questions you are going to ask the participants prior to the sessions (e.g., background questionnaire), after each task scenario is completed (ease and satisfaction questions about the task), and overall ease, satisfaction and likelihood to use/recommend questions when the sessions is completed.

- **Quantitative metrics**: Indicate the quantitative data you will be measuring in your test (e.g., successful completion rates, error rates, time on task).

- **Roles**: Include a list of the staff who will participate in the usability testing and what role each will play. The usability specialist should be the facilitator of the sessions. The usability team may also provide the primary note-taker. Other team members should be expected to participate as observers and, perhaps, as note-takers.

Identifying Test Metrics
There are several metrics that you may want to collect during the course of testing.

- **Successful Task Completion**: Each scenario requires the participant to obtain specific data that would be used in a typical task. The scenario is successfully completed when the participant indicates they have found the answer or completed the task goal. In some cases, you may want give participants multiple-choice questions. Remember to include the questions and answers in the test plan and provide them to note-takers and observers.

- **Critical Errors**: Critical errors are deviations at completion from the targets of the scenario. For example, reporting the wrong data value due to the participant's workflow. Essentially the participant will not be able to finish the task. Participant may or may not be aware that the task goal is incorrect or incomplete.

- **Non-Critical Errors**: Non-critical errors are errors that are recovered by the participant and do not result in the participant's ability to successfully complete the task. These errors result in the task being completed less efficiently. For example, exploratory behaviors such as opening the wrong navigation menu item or using a control incorrectly are non-critical errors.

- **Error-Free Rate**: Error-free rate is the percentage of test participants who complete the task without any errors (critical or non-critical errors).

- **Time On Task**: The amount of time it takes the participant to complete the task.

- **Subjective Measures**: These evaluations are self-reported participant ratings for satisfaction, ease of use, ease of finding information, etc. where participants rate the measure on a 5 to 7-point Likert scale.

- **Likes, Dislikes and Recommendations**: Participants provide what they liked most about the site, what they liked least about the site, and recommendations for improving the site.

Recruiting Usability Test Participants

It is vital to recruit participants who are similar to your site users for your usability testing. Depending on the site or product, you may have multiple potential users groups. Try to include representatives of all these groups or optimally, perform testing with each group separately if you really want to concentrate on role-based information or functionality.

Remember that you are not your user. Use your internal staff as participants only if…

- They have had no involvement in the design or development of the site or product and

- They represent a target audience

Internal staff may be used for pilot testing since you are testing the technology and the flow of the test and the data is not factored into the final results.

Internal staff should never be used to supplement during testing. It is not just any participants you need; it's participants from the right audience.

How many participants are enough?

Nielsen outlines the [number of participants](#) 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.

- **Card sorting**: test at least **15 users**.

- **Eye tracking**: test **39 users** if you want stable heat maps.

If you are testing in the federal space, please review OMB guidelines related to the [Paperwork Reduction Act for usability testing](#). For diagnostic usability testing, six to eight users of a given target audience are usually enough to uncover the major problems in a product.

Note: If you plan to do iterative (repeated) usability testing over the course of developing the site, you will need to recruit a new group for each test. You will need to factor that into your planning, recruitment, and budgeting.

Screening Participants

Participant screeners are composed of questions that will help those recruiting for your test to rule individuals in or out of contention. They may be as simple as gender and age or as complex as your target audience dictates.

Scenarios

Scenarios describe the stories and context behind why a specific user or user group comes to your site.! They note the goals and questions to be achieved and sometimes define the possibilities of how the user(s) can achieve them on the site.

Scenarios are critical both for designing an interface and for usability testing.

What to Consider When Writing Scenarios

Good scenarios are concise but answer the following key questions:

- **Who is the user?** Use the personas that have been developed to reflect the real, major user groups coming to your site.

- **Why does the user come to the site?** Note what motivates the user to come to the site and their expectations upon arrival, if any.

- **What goals does he/she have**? Through task analysis, you can better understand the what the user wants on your site and therefore what the site must have for them to leave satisfied.!

Some scenarios also answer:

- **How can the user achieve their goals on the site?** Define how the user can achieve his/ her goal on the site, identifying the various possibilities and any potential barriers.

Types of Scenarios

Types	Examples
Goal- or Task-Based Scenarios state only what the user wants to do.! Do not include any information on how the user would complete the scenario.! These scenarios are useful in helping to define your site architecture and content. You should give these types of scenarios to users in a usability test. It gives them a reason and a goal for going to the site, but it lets them show you how they would use the site to accomplish that goal.	• **Example**: A parent is worried about a ten-year old refusing to drink milk and wants to know if it really makes a difference that the child is getting very little calcium. • **Example**: You are traveling to Seattle for your job next week and you want to check on the amount you can be reimbursed for meals and other expenses.

Using Scenarios in Website Design

It is impossible to write down every scenario that every user has for visiting your website. Instead, before you start to put the site together, write down 10 to 30 of the most common reasons that users have for visiting or tasks that users want to do.

Scenarios can also work together with personas by serving as the stories behind why the particular persona would come to your website. What does the persona hope to accomplish by visiting the website? What characteristics of the persona might help or hinder his or her site interaction?

You should focus on users and their tasks rather than on your site's organization and internal structure. As a result, you will know what content the site must have and how it should be organized.

Using Scenarios in Usability Testing

When identifying scenarios for usability testing, you should limit your test to 10 to 12 tasks due to time constraints. Additionally, in a usability test, you can ask users for their own scenarios. Why would they come to your site? What do they want to do?

Usability testing scenarios should not include any information about how to accomplish a task. The usability test will show how the participant accomplishes a task and shows you whether the interface facilitates completing the scenario.

You should, however, write down how to accomplish the task. This information is included in the material that the observers and note-takers will use. Include the main pathway and any alternative pathways the participant may use to accomplish the scenario. After the test, compare how you thought users would complete the task to how they actually completed the task. This comparison provides valuable insight into the effectiveness of your site's architecture and navigation.

Running a Usability Test

Once you have [planned](#) [your test](#) and [recruited your test participants](#), 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](#), 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

Pilot Testing

Prior to conducting a usability test, make sure you have all of your materials, consents and documentation prepared and checked. It is important to pilot test equipment and materials with a volunteer participant. Run the pilot test 1-2 days prior to the first test session so that you have time to deal with any technical issues, or change the scenarios or other materials if necessary. The pilot test 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

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 Web site.
- 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.

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 behaviors, 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.

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. 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 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.