4.2: Usability Test Methods

Learning Goals

* Compare test methods for conducting evaluative research
* Discuss various environments and methodologies for conducting usability testing

 Estimated Read Time: 1 Hour.

Introduction

Welcome back! In the previous Exercise, we established that **utility + usability = useful**. For something to be useful, it must address the functional and emotional requirements expressed by your target audience during generative research. In other words, it has to be useful for the user persona or personas you identified in your initial research.

At this point, you’ve conducted interviews and card sorts to explore user needs and decided on the features and functionality needed to accomplish specific goals. You’ve also built your initial draft of these features in the form of prototypes. These features provide the **utility** necessary for something to be **useful**. Assuming you’ve successfully determined and built the features you need through your generative research methods, we can now isolate the other important variable in the equation—**usability**.

**Usability** refers to the ease with which people can use your product or experience. In order to test a product’s usability, we conduct usability tests. **Usability testing** refers to a collection of methods that design researchers use to measure and improve usability. Usability testing comes in all shapes and sizes. There’s no definitive checklist stating which methods are the most effective as the effectiveness depends on the goals of your business, your particular project, and your customers. Having said that, all methods do have a few things in common:

* Test participants that represent a subset of the audience you're designing for
* A realistic set of tasks for participants to attempt
* Users’ actions and thoughts recorded as data

In this Exercise, we’ll be exploring the ins-and-outs of usability testing from its various types to its methods of conduction.

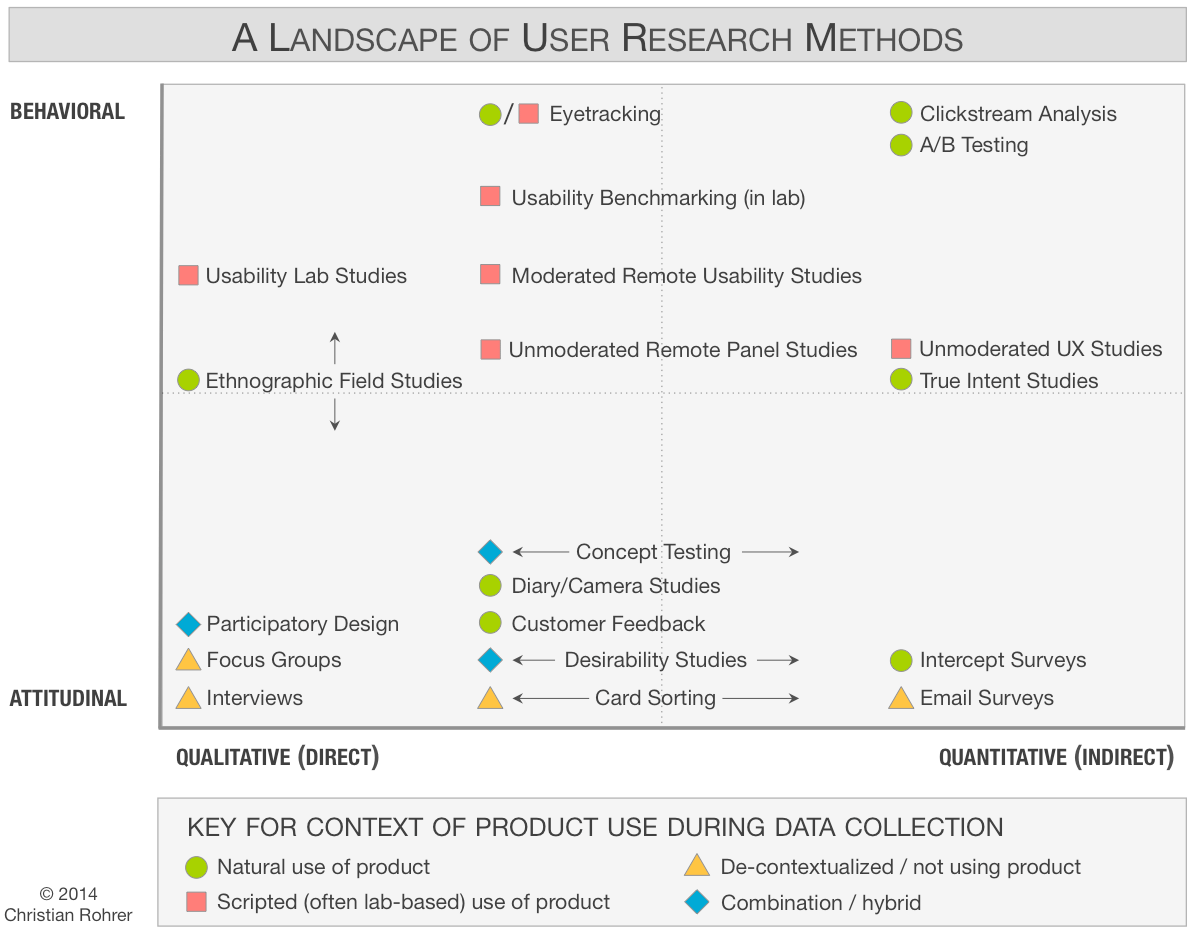
TIP!  
It’s easy to get all these “u-words” mixed up. If you get lost during the Exercise, scroll up and check out Jakob Nielsen’s succinct definition and equation as described in "[Usability 101: Introduction to Usability](https://www.nngroup.com/articles/usability-101-introduction-to-usability/)":

* **Utility** = whether it provides the features you need.
* **Usability** = how easy and pleasant these features are to use.
* **Useful = usability + utility.**

Evaluative Research Methods

In previous Achievements, we explored **generative research methods** such as interviews and open card sorts to better understand users’ attitudes and mental models. This type of research helped us identify real pain points, and, thus, real problems to solve. Now that you have a working hypothesis and something to evaluate (your prototype), let’s take a look at the other half of the research spectrum, **evaluative research methods**.

You may recall from UX Fundamentals that Christian Rohrer, leader of UX teams at Yahoo!, eBay, and Capital One, identified 20 different user research methods for UX designers to choose from. It’s worth noting that some methods are used more often than others, and it’s unlikely that any one project would utilize all of these methods. The most common methods involve some form of **usability testing**, which we’ll discuss later in the Exercise.



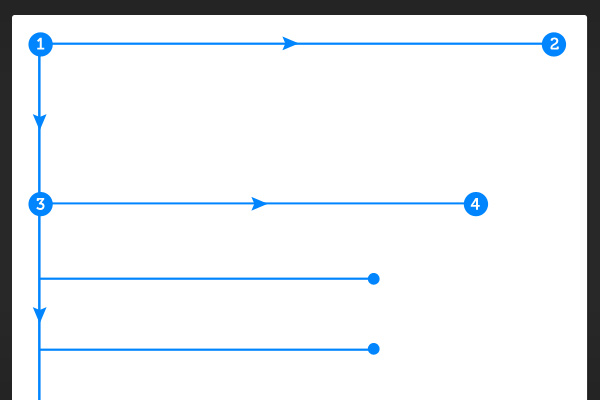
Source: [Christian Rohrer](https://www.nngroup.com/articles/which-ux-research-methods/)

Do you remember this graph from Achievement 2? The User Research Methods Landscape places different methods of research in different quadrants on the graph. Its four opposing dimensions comprise behavioral (what people do) versus attitudinal (what people think/say) and qualitative (descriptive data) versus quantitative (numerical data). Whether you choose a method that’s more quantitative than qualitative or more behavioral than attitudinal (or vice-versa) will depend on the information you’re trying to obtain. You may even use more than one method. Quantitative data, for example, is incredibly useful in determining the magnitude and degree of behaviors and attitudes, but it doesn’t always give us insight into the “why” or how to fix a problem. Likewise, there are stages when researching “what people say” is more important than researching “what people do” (and vice versa). Everything depends on the situation at hand.

**Context of use** is the third dimension of the graph, referring to the context in which we gather feedback from people. Context of use can be incredibly important depending on the nature of the problem you’re trying to solve. If you’re designing technology to help engineers install satellites on top of skyscrapers, for example, the context of use might be critical as it’s a highly unique circumstance. In such cases, it might be useful to observe natural use to better understand the environment.

It’s unlikely that you’ll employ some of the below techniques with your project due to cost and lack of facilities/resources, but it’s worth defining some of the more common evaluative research methods you may employ in your future role as a designer. Let’s take a quick look at eye tracking, ethnographic field studies, and A/B testing before moving onto usability testing (usability lab studies).

Eye Tracking



**Eye tracking**, or “gaze tracking,” is the process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head. The image above shows how a user typically scans a webpage. An eye tracker is a device for measuring eye positions and eye movement and can be configured to measure precisely where participants look as they perform tasks or interact naturally with websites, applications, physical products, or environments.

* **Pros:** Eye tracking can be useful in determining the location of specific elements. The goal is to understand people’s eye gaze patterns so we can guide their eyes in the right direction or strategically place important elements in their natural line of sight.
* **Cons:** Eye tracking equipment can be incredibly complex and difficult to calibrate. It’s possible to use remote versions to track mouse movement with the hypothesis that mouse movement closely correlates to eye movement, but it can be difficult to glean accurate insights from a large group.
* **When to Use:** Sophisticated organizations with adequate funding and expertise can employ eye tracking as additional data to inform their design decisions. Heatmaps that aggregate eye movement or mouse movement can provide a useful snapshot of activity to help inform design.

Ethnographic Field Studies

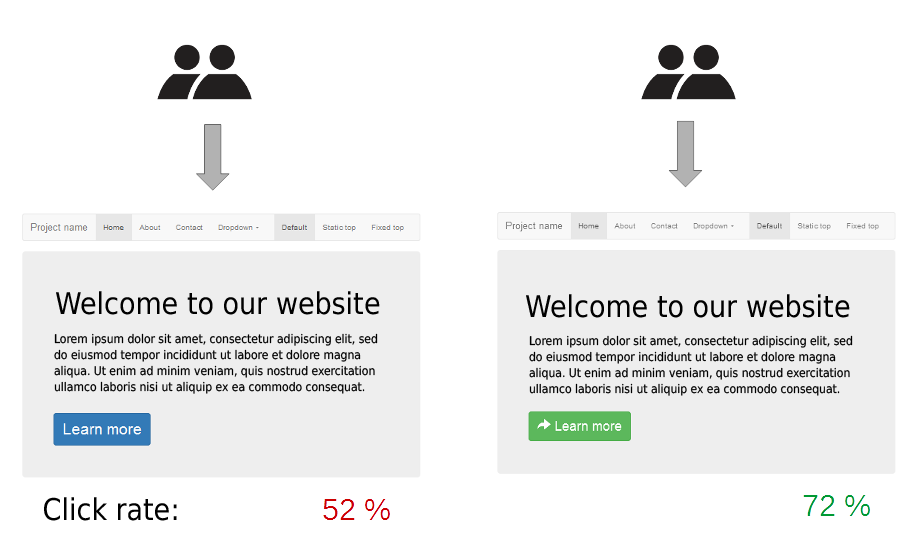
**Ethnographic field studies** involve observing participants in their natural environment. You might notice they sound similar to guerrilla testing, which we discussed in the previous Exercise; however, guerrilla testing tends to focus on the functionality of the system in question, while ethnographic field studies tend to focus on the test participants’ personal attitudes, behaviors, and environments.

* **Pros:** These studies can yield incredibly rich information about not only behavior and attitudes with a product or service, but a participant’s general habits and environment, as well. Given the time and funding, it can yield valuable information about your customers that you’d miss out on by focusing simply on the product and its usability.
* **Cons:** Tests can be expensive from both a time and a financial standpoint.
* **When to Use:** Given sufficient time and funding, these studies can be a great step towards designing more accurately for highly specific contexts and use cases.

A/B Testing

**A/B testing**, also known as “live testing,” “bucket testing,” or “split testing,” is a method of scientifically testing different designs on a site by randomly assigning groups of users to interact with each of the different designs and measuring the effect of these assignments on user behavior. You could test, for example, multiple versions of a sign-up process for your site, multiple email layouts for your newsletter, or the success of text versus video in an online tutorial. When more than two choices are necessary, researchers might also use a similar process called “multivariate testing.”

* **Pros:** Quantitative results provide definitive direction so long as the test setup and results analysis are sound.
* **Cons:** There must be significant traffic and usage of your product or service in order to garner useful data, and it can take a long time to gather enough data to reach statistically significant results. Expertise in data science is also required to make an accurate decision as to statistical significance.
* **When to Use:** For products or services with significant traffic and usage that also have teams with expertise in setting up A/B Tests and downloading results.



Source: [Maxime Lorant](https://commons.wikimedia.org/wiki/File:A-B_testing_example.png)

5-Second Tests

Yes, it really only takes five seconds! Also referred to as memory tests, **5-second tests**are used to determine whether a particular webpage or app screen does a good job of getting its core message across. For example, if you have a webpage of which the main purpose is to have visitors sign up for a newsletter, you could use a 5-second test to determine whether the users recognized that this page was asking them to sign up for a newsletter.

* **Pros:** A 5-second test is quick and easy to run. You just need a representation of the webpage or app screen you’d like to test—which can be anything from a paper prototype to a high-fidelity mockup—and a set of questions you’d like to ask users about what they remember from looking at the screen. Users are then shown the screen for five seconds and asked the prepared set of questions before the screen is covered or switched to a different one.
* **Cons:** 5-second tests aren’t well suited for screens that have multiple functions as they will be difficult for users to assess in such a short space of time, and different users are likely to focus on different function's features, leading to inconclusive results.
* **When to use:** 5-second tests are best suited to testing screens that have a single primary purpose such as signing up for a service or downloading a piece of software.

Here is a nice article on 5-second tests that's worth checking out. Although it is a bit dated (2007), it still provides a very nice and relevant explanation of 5-second tests: [5-Second Tests: Measuring Your Site’s Content Pages](https://articles.uie.com/five_second_test/).

TRY IT!  
To conduct your very own 5-Second Test, follow this [UsabilityHub link](https://fivesecondtest.com/) for more details.

We’ll carry out a form of A/B testing called Preference Testing later on in this Achievement. For now, let’s dig into our last important evaluative research method—usability testing.

All About Usability Testing



As we mentioned earlier, usability testing refers to a collection of methods that design researchers use to measure and improve usability. It’s a type of evaluative research used to measure the usability of your design by testing an interactive version of the design with users. In UX Fundamentals, you conducted a basic usability test of your mobile app design with a few users. Now, let’s take a moment to review the finer points of usability testing and why it’s so important.

What Is Usability Testing?

**Usability testing** is a technique used by UX designers, design researchers, and interaction designers to evaluate the utility and usability of a prototype or live product. Due to the flexible nature of usability testing, it’s the most common method employed by UX designers, which is why we’ll be focusing on it for this course.

As the name implies, usability testing involves testing the usability of your product on a set of users. It’s a great way to evaluate what people say and do when interacting directly with your product or service. Most often, users are given a specific set of tasks to complete, and the efficiency, ease, and speed with which they can complete them are recorded. Keep in mind that it’s not the participants themselves being tested but the effectiveness of the product or service you’ve designed.

Why Conduct Usability Tests?

An effective usability test reveals how a typical user would attempt to complete a task using your product. It can also grant insights into how to fix problems that may arise during the test. Researchers can observe where participants get confused or make wrong choices, take notes, and then implement steps to correct the issues. Usability testing doesn’t have to be expensive, and it can be a great way to quickly gain insights that inform design decisions. A well-structured and successfully executed usability test helps designers make better decisions by demonstrating:

* The ability of participants to complete tasks
* How long it takes participants to complete tasks
* Whether participants enjoy interacting with the product or service
* Whether the product or prototype has any critical errors

As we saw in UX Fundamentals, conducting this type of test before rolling out a final product design saves time and money for a company and helps ensure the final product will be well-received by its intended users. Rather than spend resources creating a final, polished product, designers can quickly iterate and course-correct several times before investing in a final design to hand off to developers.

Who Is Involved?

Ideally, you’d have at least two researchers involved—one to conduct the usability test and another to take notes and record results.

In the previous Exercise, we discussed “who” you should test with, as well as how to recruit participants. The profile of your participants will depend on the nature of the test itself. If you’re testing purely for raw functionality (whether someone can complete a specific task), the accuracy of the target audience isn’t critical. In this case, you can be more flexible with the profile of your participants. If, however, you’re trying to determine the attitudes and behaviors of a very specific audience (ER doctors or satellite engineers, for example), you may need to recruit highly specific and specialized participants, which tends to cost more from both time and financial standpoints.

Where Do Usability Tests Take Place?

Test environments vary according to the scope and nature of your project. There’s a common misconception that usability testing requires a formal usability lab and that recruiting participants is expensive. In fact, there are a variety of ways to conduct an effective usability test depending on the nature and constraints of the project. Ultimately, the test environment and setting will depend on the goals of your study, as well as the time and geographical constraints on the researcher and test participants. Try asking yourself the following questions about your project for a better idea of what a proper usability test will require:

* Does the usability test require a moderator to facilitate the test?
* Will the test take place remotely or in-person?

These considerations will determine whether your usability test will take place in a lab or out in the real world, otherwise known as “the field” in research terminology. Here’s a look at the variety of possibilities for test environments:

* **Guerrilla Testing**, or “Hallway Testing,” as you’ll recall, involves intercepting people in their natural environment and asking them to complete quick tests. This is especially effective if you have access to your target audience and the nature of your study doesn't require sensitive information from participants. It’s all right to ask, for example, “Can you add this T-shirt to your cart?” as the primary subject is the T-shirt and the capability of the system. In contrast, it might not be a good idea to walk up to someone and ask “When’s the last time you’ve taken prescription medication?” This is about the participant’s personal situation and should be handled with informed consent and additional explanation, making it a difficult subject to broach with a guerrilla test.
* **Field Studies** involve observing participants in their natural environment—where they’d be most likely to engage with your product or service. While guerrilla testing tends to focus on the functionality of the system in question, field studies tend to focus on the test participant’s personal attitudes, behaviors, and environments.
* **Usability Test Labs** require an investment in time, space, and equipment, but can be a great way to facilitate tests that allow for observations from other researchers and team members. Many companies and academic institutions have built Usability Labs that allow researchers to bring participants in for interviews, then test and record participants in a seemingly natural yet highly sophisticated setting. Usability Test Labs often have a variety of devices such as audio and recording equipment and sometimes even a two-way mirror to allow stakeholders and researchers to observe without distracting the participant.

When Do Researchers Conduct Usability Tests?

This is probably the hardest question for designers to answer. Not only does it depend largely on the project itself, it also requires experience on the part of the designer—enough projects under their belt that they’ve developed a “feeling” for when and what to test. Usability tests are an effective way of obtaining feedback and should be conducted throughout the design process, but it’s ultimately up to the design researchers to determine the best time and factors to test to achieve the best results.

Every project is different. Good designers know *how* to conduct usability tests. Great designers also know *when* to test. At this point, focus on *how* to test. Once you’re working as a designer in the field, you’ll be able to observe *when* usability tests yield the best results for each particular project.

How Much Do Usability Tests Cost?

Short answer? It depends, just like so much when it comes to usability tests. Costs can vary wildly depending on the complexity, budget, and depth of research necessary to make good design decisions. Where you’re conducting your research can also play a factor in the cost, so it’s difficult to give exact numbers. In the U.S., rates can range from free all the way up to hundreds of dollars per response if utilizing third parties or if the participant is of a highly specific target audience.

Here are a few factors that can influence the cost of your research:

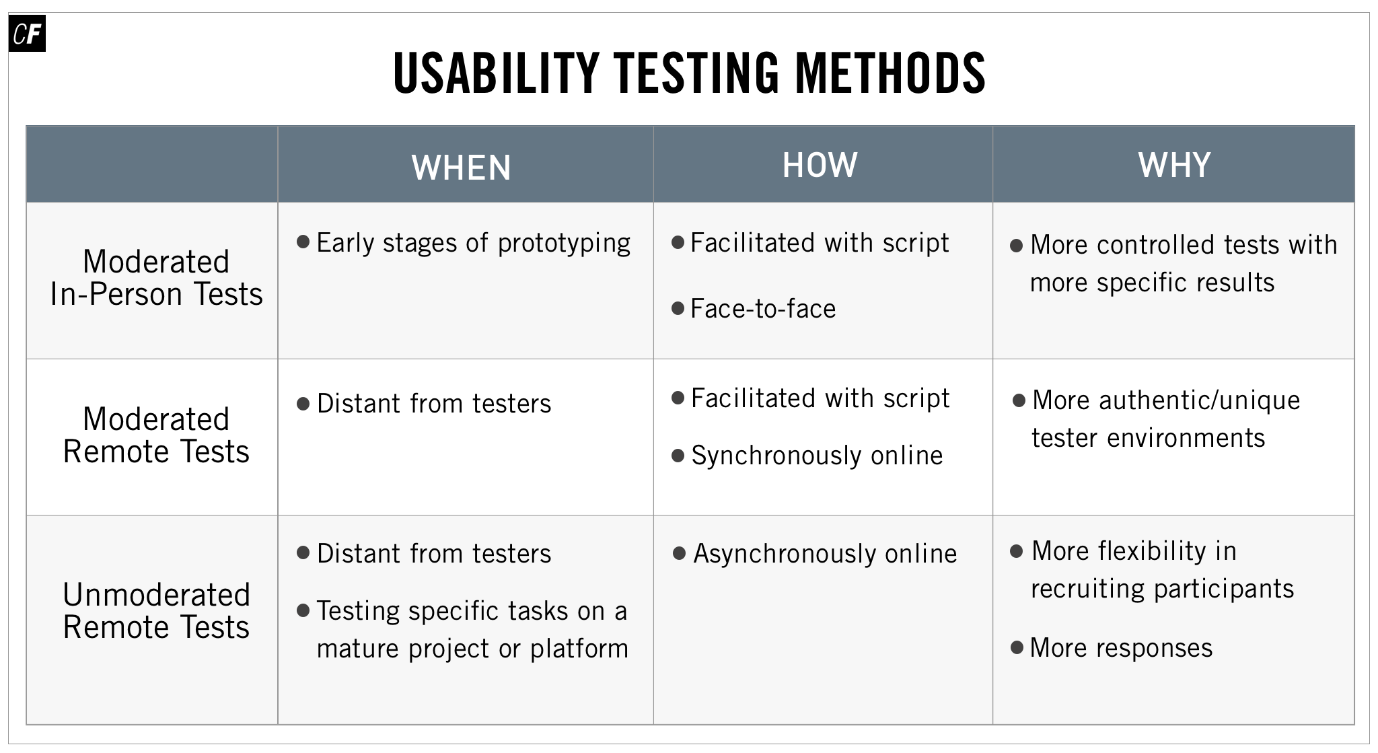
* **Usability test method.** (exact methods will be covered in the next section)
* **Number of participants.** Oftentimes, you only need five participants to catch most usability issues and inform design decisions. If you need to conduct statistically significant studies for more scientifically precise projects, however, you may need to conduct the same test with twenty or more participants.
* **Time allocated.** Keep in mind that for every hour of testing, you’ll probably need two hours afterwards to download results and communicate findings to the rest of your team. If possible, it’s best to run multiple short tests instead of long, time-consuming tests. Time is money, as they say.

How to Conduct Usability Tests

Ideally, usability tests are conducted with test participants from a specific target audience. Participants are usually asked to complete tasks using the product or prototype while the researchers watch, listen, and take notes.

It’s important to have a test plan and test script to make sure the tests are consistent. We’ll be working on creating these over the next couple of Exercises. The primary goal of your tests is to pinpoint usability problems through behavioral observation and determine participants’ satisfaction with the product or prototype.

You may remember from the Fundamentals Course that there are three basic types of usability test methods researchers use to gather feedback from real people: moderated in-person, moderated remote, and unmoderated remote. Let’s take another look at them in-depth.



[Click here](https://s3.amazonaws.com/coach-courses-us/public/courses/ux-immersion/A4/4.2/A4E2_usabilitytestmethods.png) to zoom in

Moderated In-Person Tests

**Moderated in-person tests** require a researcher or facilitator co-located with the participant (often in a lab, though locations can vary). The researcher and/or observer facilitates the test via a script, typically creating a more controlled test with more specific results. This is because a moderator is able to ask follow-up questions and can observe body language in addition to verbal responses. There are, however, a few drawbacks to this method. First, this kind of testing tends to be more time-consuming as the moderator must be physically present, and second, that same physical presence can change the way certain participants respond, potentially skewing data.

**When to Use:** Moderated in-person tests are especially useful during the early stages of prototyping. At this stage, the prototype is likely missing certain functionality, and users might require guidance to complete the desired tasks. Moderated tests are particularly effective at digging deeper into the “why” behind a participant's behavior, which can be crucial in guiding design direction and feature priority at the beginning of a project.

Alternatively, if you have a very mature and complex prototype, a moderator can help keep test participants on track—only testing the current area and tasks of interest. While recruiting participants can be time- and cost-intensive, researchers often manage to find local participants to volunteer their time for little to no cost, especially if interest already exists for the product or service.

Moderated Remote Tests

**Moderated remote tests** are useful when the participant and facilitator are in different locations. Screen sharing software such as GoToMeeting, Skype, and Slack, among others, allow the facilitator to watch remotely as the participant attempts to complete specific tasks using their product or website.

**When to Use:** Moderated remote tests are useful when participants and facilitators are geographically distant and can’t meet in person. It allows for participants to use their own technology in their own environment, which can sometimes yield more authentic results.

Unmoderated Remote Tests

**Unmoderated remote tests** don’t require the physical presence of a moderator and are, instead, conducted asynchronously online using a variety of specialized tools such as [UserTesting](http://usertesting.com/), [UserZoom](http://userzoom.com/), or [UsabilityHub](http://usabilityhub.com/). The asynchronous nature of online tests gives researchers more flexibility in recruiting participants and running tests, allowing for a higher volume of responses. One drawback, however, is that researchers aren’t able to observe body language or ask follow-up questions.

Gathering mountains of unmoderated data can be valuable so long as you have the time, resources, and knowledge to analyze the results and communicate them effectively to the rest of your team. Don’t pay for a thousand hours of recordings of your website, for example, if you don’t have a thousand hours of time to devote to watching the videos and creating useful conclusions. If your time is limited, it might be better to aggregate results in the form of a heatmap.

**When to Use:** Unmoderated tests are most useful when testing very specific tasks for a mature project or platform. Increased granularity and a higher volume of responses can be useful when evaluating a specific task or feature. Unmoderated tests also mitigate geographic and scheduling differences between researchers and participants. Online tools can be expensive, so a certain level of funding will likely be required to conduct sophisticated unmoderated tests.

What’s true of the evaluative test methods as a whole is true of the ways in which you can conduct usability tests (moderated in-person, moderated remote, unmoderated remote)—they each come with their own advantages and disadvantages, and choosing one will depend on a variety of factors, including your budget, time, and testing purpose.

Summary

Evaluative research provides an important counterpoint to generative research. While the generative research you conducted towards the beginning of your project helped shape your persona, the problem you’re addressing, and the features you’ll design, evaluative research is your chance to test whether those features are usable with real people (and update your design accordingly). There are many methods that fall on the evaluative research side of the spectrum, including advanced methods such as eye tracking, field studies, and A/B testing. The most common and important way to conduct evaluative research, however, is via usability testing.

Usability tests are an easy way for real users to interact with your prototype and for you to observe what they do and where they struggle. It’s also fairly flexible and can be done in a number of ways, from moderated to unmoderated and even remote testing. Throughout the rest of this Achievement, we’ll be covering how to set up your test plan, test script, and test environment to run a successful usability test. See you there!

Resources

* [Testing Guides from UsabilityHub](https://usabilityhub.com/guides?utm_source=UsabilityHub+Customers&utm_campaign=4100badacc-Monthly_Email_May_2017&utm_medium=email&utm_term=0_23bdfd0b30-4100badacc-171640813)
* [Example Tests from UsabilityHub](https://usabilityhub.com/examples?utm_source=UsabilityHub+Customers&utm_campaign=4100badacc-Monthly_Email_May_2017&utm_medium=email&utm_term=0_23bdfd0b30-4100badacc-171640813)
* [When to Use Which User-Experience Research Methods](https://www.nngroup.com/articles/which-ux-research-methods/)
* [Generative vs. Evaluative Research: What’s the Difference and Why Do We Need Each?](https://www.usertesting.com/blog/2015/12/17/generative-vs-evaluative-research/)
* [5 Types of Usability Tests](https://measuringu.com/five-types-usability/)
* [First Rule of Usability? Don’t Listen to Users](https://www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/)
* [Consider These Variables Before You Choose a UX Research Method](https://www.usertesting.com/blog/2016/04/22/choose-research-method/)
* [F-Shaped Pattern for Reading Web Content](https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/)

Take the quiz to test your knowledge on this Exercise.

Take Quiz

Task

* [DIRECTIONS](https://careerfoundry.com/en/steps/test-methods#directions)
* [SUBMISSION HISTORY](https://careerfoundry.com/en/steps/test-methods#step_submission_history)

 Estimated Task Time: 2 Hours.

It’s important to have a clear understanding of how you will conduct your usability tests. In this Task, you’ll be thinking through the advantages and possible challenges of conducting your usability tests.

**Directions**

1. In order for you to gain experience of running face-to-face testing sessions, you will be conducting in-person, moderated usability testing to test the core features of your app. However, as a UX designer you will need to be able to select the most appropriate testing methodology for your projects. Thinking about your current project, briefly list the benefits and challenges of using these testing methodologies: moderated in-person, moderated remote, and unmoderated. Then, suggest possible solutions or workarounds that could help you overcome the challenges or more effectively utilize the benefits of each methodology.
2. Upload your response as a PDF file here. Feel free to share additional thoughts or ask questions in the submission box.

**Bonus Task**

Sign up for [UsabilityHub](http://www.usabilityhub.com/) and take 10 tests as a participant. This will give you an idea of how others have tested their own products or services. Keep in mind that the quality of the questions and tests will depend on the knowledge of the author, so feel free to point out ways they could have improved in their methodology.

TIP!  
If you haven’t finished recruiting participants via the email template you created in the previous Exercise, now is a good time to continue reaching out to people. You’ll be conducting a usability test in Exercise 4.5, so planning your schedule (and the schedule of your participants) now will help you begin testing on time.