



World Leaders in Research-Based User Experience

Quantitative vs. Qualitative Usability Testing



Raluca Budiu

October 1, 2017

Summary: Both these complementary types of user research play important roles in an iterative design cycle. Qualitative research informs the design process; quantitative research provides a basis for benchmarking programs and ROI calculations.

All usability-testing studies involve a participant performing some assigned tasks on one or more designs. There are, however, two types of data that can be collected in a user-testing study:

- **Qualitative (qual) data**, consisting of observational findings that identify design features easy or hard to use
- **Quantitative (quant) data**, in form of one or more metrics (such as task completion rates or task times) that reflect whether the tasks were easy to perform

In This Article:

[Qual Research](#)

Qual Research

Qualitative data offer a direct assessment of the usability of a system: researchers will observe participants struggle with specific UI elements and infer which aspects of the design are problematic and which work well. They can always ask participants followup questions and change the course of the study to get insights into the specific issue that the participant experiences. Then, based on their own UX knowledge and possibly on observing other participants encounter (or not) the same difficulty, researchers will determine whether the respective UI element is indeed poorly designed.

Quant Research

Quantitative data offer an indirect assessment of the usability of a design. They can be based on users' **performance** on a given task (e.g., task-completion times, success rates, number of errors) or can reflect participants' **perception of usability** (e.g., satisfaction ratings). Quantitative metrics are simply numbers, and as such, they can be hard to interpret in the absence of a reference point. For example, if 60% of the participants in a study were able to complete a task, is that good or bad? It's hard to say in the absolute. That is why many quant studies usually aim not so much to describe the usability of a site, but rather to compare it with a known standard or with the usability of a competitor or a previous design.

While quant data can tell us that our design may not be usable relative to a reference point, they do not point out what problems users encountered. Even worse, they don't tell us what changes to

make in the design to get a better result next time. Knowing that only 40% of the participants are able to complete a task doesn't say why users had trouble with that task or how to make it easier. Often researchers will need to use qual methods to supplement quant data in order to understand the specific usability issues in an interface.

Statistical Significance

One advantage of quant over qual is **statistical significance**. When quant data are presented in a sound way, they come with a certain protection against randomness: usually, mathematical instruments such as confidence intervals and statistical significance will tell us how likely it is that the data reflect the truth or whether they may be just an effect of random noise — perhaps an artifact of the specific participants that we happened to recruit or of the conditions in which the study was run. Although seasoned qual researchers will deploy an arsenal of good practices to protect themselves from chance and to make sure that their results are not biased, we have no formal assurance that the findings from a qual study are indeed objective and representative for the whole target population.

Differences Between Qual and Quant

Qualitative and quantitative data need slightly different study setups and very different analysis methods. They are rarely collected at the same time — hence the distinction between qualitative and quantitative user studies. Both qualitative and quantitative testing are essential in the iterative design cycle. Although qual studies are more common in our industry, quant studies are the only ones that allow us to put a number on a redesign and clearly say how much our new version improved over the old one — they are the essential instrument in calculating [return on investment](#).

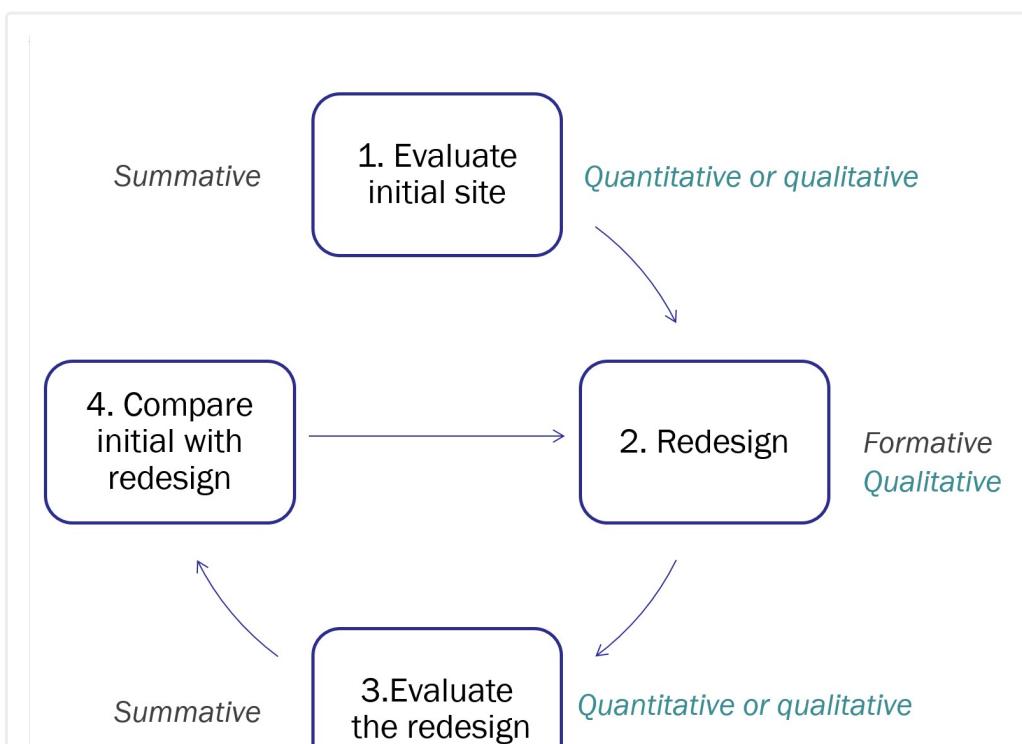
The table below summarizes the differences between the two types of research. In the rest of the article we discuss these differences in detail.

	Qual Research	Quant Research
Questions answered	Why?	How many and how much?
Goals	Both formative and summative: • inform design decisions	Mostly summative: <ul style="list-style-type: none">• evaluate the usability of an existing site

	<ul style="list-style-type: none"> • identify usability issues and find solutions for them 	<ul style="list-style-type: none"> • track usability over time • compare site with competitors • compute ROI
When it is used	Anytime: during redesign, or when you have a final working product	When you have a working product (either at the beginning or end of a design cycle)
Outcome	Findings based on the researcher's impressions, interpretations, and prior knowledge	Statistically meaningful results that are likely to be replicated in a different study
Methodology	<ul style="list-style-type: none"> • Few participants • Flexible study conditions that can be adjusted according to the team's needs • Think-aloud protocol 	<ul style="list-style-type: none"> • Many participants • Well-defined, strictly controlled study conditions • Usually no think-aloud

The Iterative Design Cycle: Goals for Qual vs. Quant

The basic user-centered design cycle starts with an evaluation of an existing design, followed by a redesign intended to address the current system's usability challenges. Once the new version is complete, it can be evaluated and compared against the initial version.



The iterative design cycle: Steps 1 and 3 involve summative research (done with either quantitative or qualitative methods), while step 2 involves formative research (done with qualitative methods).

The first and third stage of the iterative design cycle are **summative** — they are intended to provide an overall assessment of a design. In these steps both qual and quant research methods (or combinations such as [PURE](#)) can be used to evaluate the design. However, when the goal is to link the entire redesign effort to actual financial savings or explicitly figure out how much the redesign has improved, quant studies must be used. Organizations with mature UX often have such a quant usability-tracking process in place. (Sometimes this process of quantitatively evaluating each version of a design and comparing it with previous versions is called **benchmarking**.)

During the redesign stage, user research has a **formative** role: it is meant to inform the design and steer it on the right path. In this phase, designers and researchers need to get user data relatively quickly and cheaply in order to be able to choose among different design alternatives and create a usable UI. At this stage, **qual studies are usually the most appropriate**. We know that, [with 5 users, a qualitative study is likely to uncover 85% of the usability problems](#) in a design (provided that the design is not already close to being perfect), so, in the redesign step, it makes sense to run one quick study with a few users, determine the big issues, fix them, and then test the new version again with another small set of users.

When to Use Qual vs. Quant

Qual studies are well suited for identifying the main problems in a design: for example, we can easily run a qualitative study to see what (if anything) prevents users from submitting a [form](#) successfully, and, based on that study, we may determine that we need to lengthen the form fields, present [password requirements](#), or use [labels outside the fields](#) rather than inside them.

In contrast, most quant studies are done on a complete version of the site, with the purpose of **evaluating the usability of the site**, rather than directly informing the redesign process. This is not because one could not employ quantitative methods during the redesign iterations, but rather because quant usability studies would be too costly if used often and early in the design process. Quant studies usually involve a [large number of users](#), and most organizations cannot afford to spend a lot of money on such studies to investigate whether the page copy is clear or whether a button is findable. However, the numbers obtained from quantitative testing can be invaluable when it comes to convincing upper management that your site is in need of a complete redesign.

Outcome: Qual vs. Quant

Qual data usually will consist of a set of findings, which identify (and prioritize according to severity) the strengths and weaknesses of a design. These findings are estimates — they are based on the knowledge and level of experience of the researcher facilitating the task and interpreting the meanings of the users' actions. Different practitioners will often identify different issues in the same user-testing session (a phenomenon known as the **evaluator's effect**). Plus, even if we've been very careful in recruiting participants who match our target demographics, when we include only a few people there is always a chance that they are not truly representative of the whole user population, and so our findings may be skewed.

Quant studies usually involve a relatively large number of users (often more than 30) and use statistical techniques to protect themselves against such random events. When reported correctly, quantitative studies will include information about the **statistical significance** of the results. For example, a margin of error will help you understand how much you can trust the results from the study. Or, if the difference in task-completion time between your site and the competitor's site was statistically significant, you will know that, even if you were to recruit a different set of users and rerun your study, your results will point in the same direction, even if the exact averages may be slightly different.

Thus, when quantitative studies are carried out and analyzed correctly, you can have confidence that their results are sound. Namely, that they are not due to a lucky or unlucky throw of the dice.

These types of analyses are based on statistics and usually involve other types of skills than you will find in a qual usability researcher. That is why many companies have separate job requirements for quant and qual UX researchers.

Methodology: Qual vs. Quant

On the surface quantitative and qualitative user testing can look quite similar (i.e., they both involve users performing tasks on a design). Both types of studies need to follow the basic rules for good experiment design, by making sure that they have:

1. **External validity:** participants are representative of the target audience and the study conditions reflect how the task is done in the wild. For instance, testing a mobile site on a desktop simulator lacks external validity because people would normally use that site on a touchscreen phone.
2. **Internal validity:** the experiment setup does not favor any one condition. For example, if design A is tested in the morning and design B is tested in the afternoon, it is possible that fatigue play a role in how participants use design B.

But, because quant studies strive to obtain results that are meaningful statistically, there are some

important differences between the two types of studies:

- As discussed above, quant studies involve more users than qual studies.
- Because differences in the session setup and in participant backgrounds can increase the measurement noise and lead to larger margins of error, quant studies aim to minimize variability as much as possible. Thus:
 - The conditions in quant studies need to be stringently controlled from session to session. That is, you need to make sure that your participants are all run in pretty much the same environment as possible: you cannot have two sessions done in person, and three sessions done remotely.
 - Quant studies often start with a practice task intended to make all participants familiar with the study setup and with the site being evaluated. In this way, possible individual differences between, say, expert and novice users are ironed out, as novices get a chance to learn the interface.
 - The [think-aloud protocol](#) is the de facto method in qual studies, but is sometimes not recommended in quant studies. Researchers are split as to whether the think-aloud protocol can be soundly used in quant studies. To some degree, because some people are more talkative than others, it is likely to increase the measurement noise. As a result, many quant studies do not ask participants to think out loud.
 - Personal information such as names, addresses, or birthdates will increase the variability of the study, because different people have different data. Whereas for a qual study, [you want people to enter their own, real information](#), in a quant study everybody should have the same experience and thus should type in the exact same strings. That is why participants should be provided with a set of made-up data that they can all use. (This constraint will sometimes create backend difficulties for a live system.)
- Conversely, for qual studies, it's okay to vary the study conditions between sessions. For example, if you discover that a certain task doesn't give you the insights you need, by all means rewrite it before running your next user. Changing the task would make it invalid to average measures across users who had performed the different tasks, but in a qual study you aim for insights, not numbers, so you can take liberties that will ruin numbers (which are not your research goal anyway).
- For quant studies, tasks need to have a single well-defined answer. Thus, while a task such as "Find the phone number and the address for John Smith" may be appropriate for a qual study, it is not good for quant studies because it is hard to code success for it: if the participant finds the phone number but not the address, should that be considered a failure?

Moreover, all participants should understand the same thing when they read the task. A task such as "Research the requirements for obtaining a drone-flying permit in California" is too vague for a quant user study, as different people may understand different things by the word "research," but can be okay in qual studies if you're trying to figure out what types of

information people may be interested in.

- Whereas it is good practice to use task randomization in both qualitative and quantitative experiments, often qual studies won't be completely randomized. In quant tests, randomization ensures that the order of the tasks does not bias the results in any way.

Conclusion

Qualitative and quantitative user testing are complementary methods that serve different goals. Qual testing involves a small number of users (5–8) and directly identifies the main usability problems in an interface. It is often used formatively, to inform the design process and channel it in the right direction. Quant usability testing (or benchmarking) is based on a large number of participants (often more than 30); when analyzed and interpreted correctly, results from quant tests come with higher protection against random noise. Quant studies offer an indirect, summative evaluation of the usability of a site through metrics such as task-completion rate, task time, or satisfaction ratings and are typically used to track the usability of a system over design iterations.

For an overview of popular quantitative research methods, guidance on which to use each one, and how to calculate return on investment, check out our course [Measuring UX and ROI](#).

Related Courses

How to Interpret UX Numbers: Statistics for UX

When research data should be trusted; what statistics to use when

RESEARCH

Measuring UX and ROI

Use metrics from quantitative research to demonstrate value

RESEARCH

Remote User Research

Gather insights virtually through usability testing and other remote methods

RESEARCH

Learn More:

Qualitative vs. Quantitative UX Research



Qualitative vs. Quantitative UX Research

Kate Moran · 4 min



**Frequency ≠
Importance in
Qualitative Data**
Tanner Kohler · 3 min



**Product
Instrumentation: 3
Benefits**
Sara Ramaswamy · 4
min



**Between-Subject vs.
Within-Subject
Study Design in User
Research**
Raluca Budiu · 5 min

Related Articles:

Between-Subjects vs. Within-Subjects Study Design

Raluca Budiu · 8 min

Collecting Metrics During Qualitative Studies

Kate Moran · 3 min

Quantitative UX: Glossary

Raluca Budiu · 8 min

Risks of Quantitative Studies

Jakob Nielsen · 11 min

Net Promoter Score: What a Customer-Relations Metric Can Tell You About Your User Experience

Therese Fessenden · 7 min

Quantitative Research: Study Guide

Kate Moran · 8 min

Subscribe to Our Newsletter

Get weekly UX articles, videos, and upcoming training events straight to your inbox.

Subscribe

Follow Us

LinkedIn

Youtube

Twitter

Instagram

Podcast

Certification

What is UX Certification?

Specialties

Exams

UX Certified People

UX Training

Virtual Training

In-Person Training

Browse Courses

Course Calendar

Consulting

Expert Review

User Testing

Customized Research

Team Training

Applied Workshops

Keynote Speaking

Free Guidance

Articles & Videos

The NN/g UX Podcast

About

Why NN/g

About Us

People

Contact

Return Policy

Copyright © 1998-2023 Nielsen Norman Group, All Rights Reserved.

[Cookie Preferences](#)

/

[Cookie Declaration](#)

/

[Privacy Policy](#)