



World Leaders in Research-Based User Experience

Why 5 Participants Are Okay in a Qualitative Study, but Not in a Quantitative One

Summary: Qualitative usability testing aims to identify issues in an interface, while quantitative usability testing is meant to provide metrics that capture the behavior of your whole user population.

By [Raluca Budi](#) on July 11, 2021

Topics: [Research Methods](#)

In our quantitative-usability classes ([Measuring UX and ROI](#) and [Statistics for UX](#)) we often recommend a sizeable number of participants for quantitative studies — usually more than 30. We've said [again](#) and [again](#) that metrics collected in qualitative usability testing are often misleading and do not generalize to the general population. (There could be exceptions, but you always need to check by calculating [confidence intervals](#) and [statistical significance](#)). And, almost inevitably, the retort comes back — [Didn't Jakob Nielsen recommend 5 users for usability studies?](#) If you need more users for statistical reasons, then it certainly means that the results obtained with 5 users aren't valid, doesn't it?

This question is so frequent, that we need to address the misunderstanding.

Quantitative Usability Studies: More than 5 Participants

Participants

Quantitative usability studies are usually [summative](#) in nature: their goal is **to measure the usability of a system** (site, application, or some other product), arriving at one or more numbers. These studies attempt to get a sense of how good an interface is for its users by looking at a variety of metrics: how many users from the general population can complete one or more top tasks, how long it takes them, how many errors they make, and how satisfied they are with their experience. They usually involve collecting values for each participant, aggregating those values in summary statistics such as averages or success rates, calculating confidence intervals for those aggregates, and reporting likely ranges for the true score for the whole population. The results of such a study may indicate that the success rate for a top task for the whole population is somewhere between 75% and 90%, with a 95% confidence level and that the task time is between 2.3 and 2.6 minutes. These ranges (in effect, confidence intervals) should be fairly narrow to convey any interesting information (knowing that a success rate is between 5% and 95% is not very helpful, is it?), and they usually are narrow only if you include a large number of participants (40 or more). Hence, **the recommendation to calculate confidence intervals for all metrics collected** and not to rely on summary statistics when studies contain just a few users.

Qualitative Usability Studies: Assumptions Behind the 5-User Guideline

In contrast, qualitative user studies are mostly [formative](#): their goal is to figure out what doesn't work in a design, fix it, and then move on with a new, better version. The new version will usually also get tested, improved on, and so forth. While it is possible to have qualitative studies that have summative goals (let's see all that's wrong with our current website!), a lot of the times they simply aim to refine an existing [design iteration](#). Qualitative studies (even when they are summative) do not try to predict how many users will complete a task, nor do they attempt to figure out how many people will run into any specific usability issue. **They are meant to identify usability problems.**

In comes Jakob Nielsen's article that recommends qualitative testing with 5 users. There are three main assumptions behind that recommendation:

1. **That you are trying to identify issues in a design.** By definition, an issue is some usability problem that the user experiences while using the design.
2. **That any issue that somebody encounters is a valid one worth fixing.**
To make an [analogy for this assumption](#): if one person falls into a pothole, you know you need to fix it. You don't need 100 people to fall into it to decide it needs fixing.
3. **That the probability of someone encountering an issue is 31%**

Based on these assumptions, Jakob Nielsen and Tom Landauer built a mathematical model that shows that, by doing a qualitative test with 5 participants, you will identify 85% of the issues in an interface. And Jakob Nielsen has repeatedly argued (and justly so) that a good investment is to start with 5 people, find your 85% of the issues, fix them, then test again with another 5 people, and so on. [It's not worth trying to find all the issues in one test](#) because you'll spend too much time and money, and then you'll be sure to introduce other issues in the redesign.

Note that the "metrics" collected in the quantitative and qualitative studies are very different: in quantitative studies you're interested in how your general population will fare on measures such as task success, errors, satisfaction, and task time. **In qualitative studies, you're simply counting usability issues.** And, while there is a statistical uncertainty about any number obtained from a quantitative study (how will the average obtained from my study compare with the average of the general population), **there is absolutely no uncertainty in a qualitative study — any error identified is a legit problem that needs to be fixed.**

Questioning the Assumptions Behind the 5-User Guideline

I gave you a list of assumptions on which the 5-user guideline is based. However, you may not agree with (some of) them. I don't think there's much to argue about the first assumption, but you may bring some valid objections to the second and third one.

Does any error that someone encounters need to be fixed? One may argue that if 1000 out of a 1000 people fall into a pothole, you do need to repair it, but not if only one person out of 1000 falls into it. With a qualitative usability study, you have no guarantee (based on only the study) that an identified issue is likely to be encountered by more users than the ones that happened to come to your study. So, in that sense, the results cannot be generalized to the whole population.

Yes, if you wanted you could run a quantitative study to predict how many people in the general population are likely to encounter a particular error. And, then, yes, you could prioritize errors based on how likely they are and fix the ones with the highest priority. While that approach is certainly very sound, it's probably going to be also very wasteful — you will need to test your design with a pretty large number of users to identify its main problems, then fix them, and introduce another ones that will need to be identified and prioritized.

Instead, the qualitative approach assumes that designers will use some other means to [prioritize among different issues](#) — maybe some of them are too expensive to fix or others are related to a functionality that only few of your users are likely to use. **Qualitative user testing simply gives you a list of problems.** It is the researcher job's to prioritize among the different issues and move on.

Is the chance of encountering a problem in an interface 31%? The 31% number was based on an average across several projects run in the early 90s. It is possible that, since then, the chance of encountering an issue has changed. It's also possible that, as you're doing more design iterations and fixing more and more errors, the usability of your product is substantially better and, in fact, it's more difficult to encounter new issues.

The good news is that the chance of encountering an error in an interface is only

a parameter in Nielsen and Landauer's model. So, if you know that your interface is pretty good, you can simply insert your desired probability in that model. The number of users will be given by this equation:

$$N = \frac{\log 0.15}{\log(1-L)}$$

where L is your estimated probability of encountering an error in an interface, expressed as a decimal (i.e., 31% is entered as .31)

For example, if L is 20%, you would need 9 users to find 85% of the problems in the interface. If L is 10%, then you'd need 18 users. The more usable your interface is, the more users you need to include in the test to identify 85% of the usability problems.

However, **your real goal is not to find a particular percentage of problems, but to maximize the business value of your user-research program.** It turns out that [peak ROI](#) is fairly insensitive to variations in model parameters. If you are testing a terrible design for the very first time, your expenses will be low (it will be very easy to identify usability problems) and your gains will be high (the product will be hugely improved). Conversely, if you are researching a difficult problem, your expenses will be higher and your gains will be lower. However, the point that *maximizes* the ratio between gains and expenses (i.e., ROI) will still usually be around 5 test users, even though your study profitability will be higher for the easy study and lower for the harder study.

In general, it's a good idea to **start with 5 users, fix the errors that you find, and then slowly increase the number of users on further iterations** if you think that you've made great progress. But, in practice, you can easily get a sense of how much insight you've found with 5 users. If you feel that not much, by all means, include a few additional users. Conversely, you can [test with fewer than 5 users](#) under other circumstances, such as when you can proceed to testing the next iteration very quickly. But if you have plenty of issues that you need to work on, first fix those, then move on.

Conclusion

There is no contradiction between the 5-user guideline for qualitative user testing and the idea that you cannot trust metrics obtained from small studies, because you do not collect metrics in a qualitative study. Quantitative and qualitative user studies have different goals:

- **Quantitative** studies aim to find **metrics** that predict the behavior of the whole population; such numbers will be imprecise — and thus useless — if they are based on a small sample size.
- **Qualitative** studies aim for **insights**: to identify usability issues in an interface. Researchers must use judgment rather than numbers to prioritize these issues. (And, to hammer home the point: the 5-user guideline only applies to qualitative, not to quantitative studies.)

If your interface has already gone through many rounds of testing, you may need to include more people even in a qualitative test, as the chance of encountering a problem may be smaller than the original assumptions of the model. Still, it's good practice to start with 5 users and then increase the number if there are too few important findings.

Share this article: [Twitter](#) | [LinkedIn](#) | [Email](#)

Copyright © 1998-2022 Nielsen Norman Group, All Rights Reserved.