



World Leaders in Research-Based User Experience

Collecting Metrics During Qualitative Studies

Summary: Don't report descriptive statistics like success rates and averages unless you ran a quantitative study. Reported numbers must be qualified with statistical information such as confidence intervals or statistical significance.

By [Kate Moran](#) on June 13, 2021

I'm often asked variations of this question:

“Is it Ok to collect quantitative metrics (like task success and time on task) during qualitative usability testing studies?”

The short answer is: **Yes, as long as (1) you report those numbers anecdotally, not statistically; and (2) you are careful to specify that those numbers may not reflect the behavior of your entire user population.**

Quantitative & Qualitative Studies Are Different

In some cases, UX teams ask this question because they hope to blur the lines between quantitative and qualitative research to save time or resources.

However, [quantitative and qualitative UX studies](#) have different goals, and so they **must be structured differently.**

In **qualitative** user testing, the priority is on identifying problems or opportunities in the experience and determining how to fix them. Researchers focus on collecting insights and anecdotes from the study. As a result, [only a handful of representative users](#) are needed to run these studies. Additionally, the facilitator has some flexibility around tasks — they can be modified or tailored to fit each participant.

In **quantitative** user testing, the priority is on [collecting UX metrics](#) — numerical assessments of various aspects of the experience, such as the time it takes to perform a task. The end goal is to produce estimates for those metrics over the whole population of users.

In order for these estimates to be realistic and not random, a [relatively large number of participants](#) is needed (often more than 40). All these participants must attempt the same exact tasks — otherwise, the numbers collected from the study won't be meaningful, because they won't describe performance with the same task.

To summarize, **you shouldn't try to blend your quantitative and qualitative usability studies together**, because each study type is run differently.

Difference	Qualitative Studies	Quantitative Studies
Primary goals	Insights	Metrics
Required numbers of participants	Around 5	Around 40
Flexibility in study structure	Ability to adapt task instructions	All task instructions must be the same

Collecting and Reporting Metrics in Qualitative Studies

So, if you can't blend your quantitative and qualitative studies together, why would you want to collect metrics during a qualitative study?

would you want to collect metrics during a qualitative study :

In qualitative usability testing, collecting metrics can be helpful in [telling a story](#) as you report your results. The key is: **you have to [report individual values, not aggregated statistics](#) such as time averages or success rates.**

For example, imagine we ran a qualitative user test on a food-delivery app with 6 participants. We collected some quantitative metrics during this qualitative study: time on task and success rate. As we report findings and support them with anecdotes from the research, these metrics might help us communicate the severity of problems.

Correct: “One participant struggled to find a restaurant option nearby that she liked. She spent over 8 minutes browsing and applying different filters before she decided to place an order.”

It’s fine to report that individual value (“over 8 minutes”) when trying to highlight how much difficulty this participant encountered in the task. However, since we ran a qualitative study, we *cannot* use this data to report any averages.

Incorrect: “Participants took on average 4 minutes and 23 seconds to choose a restaurant option.”

It’s wrong to report averages in this case, because it implies that we expect that, if we tested all of our real-life users, we’d find a similar average time on task. Because we only tested with a handful of participants, we don’t have sufficient data to make that claim. (If we calculated a [confidence interval](#) for this data, it would be very wide.)

Similarly, for our task-success data, we can’t report a rate.

Incorrect: “Only 16.7% of participants were able to complete the task successfully.”

Again, this implies we tested with a larger sample than we really did and it makes the implicit claim that we’d expect to see a similar proportion in our full population of users — which we can’t make based on this data.

However, we can report the number of successful participants out of the number who attempted the task.

Correct: “Contacting customer support was very challenging for our study participants — only 1 out of the 6 participants was able to complete this task successfully.”

This approach accomplishes the goal of communicating that the task was difficult, without presenting a percentage which could be easily misunderstood.

For More Information

If UX metrics are important to you, consider running a dedicated quantitative usability test. While these can be much more expensive and time consuming than qualitative studies, they’re the best way to collect metrics. [Remote unmoderated usability testing](#) can help make this method more practical.

To learn the correct ways of analyzing quantitative UX data, check out our full-day course, [How to Interpret UX Numbers: Statistics for UX](#) .

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

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