



World Leaders in Research-Based User Experience

# Data Is More than Numbers: Why Qualitative Data Isn't Just Opinions

**Summary:** Systematically gathered qualitative data is a dependable method of understanding what users need, why problems occur, and how to solve them.

By [Page Laubheimer](#) on August 1, 2021

Topics: [Research Methods](#)

---

“Not everything that can be counted counts, and not everything that counts can be counted.”

– (Attributed to) Albert Einstein

A fairly common objection to [qualitative UX research](#) (especially from statistically literate audiences) is that [small sample sizes](#) result in anecdotal evidence or a few people's subjective assessments, rather than *data* proper. Many UXers that work in domains such as healthcare, natural science, or even just “data-driven” organizations may find that it is difficult to build buy-in to conduct small-*n* research in the first place; even if they are able to do the testing, it's often hard to build credibility about the recommendations that result from the findings.

Common objections include:

- Comparisons between design options in studies with 5 or 10 users [aren't statistically significant](#) (which is true).

- Small sample sizes mean that we cannot confidently generalize things like [time on task or success rates from a small study](#) (also true).
- Since we aren't *measuring* things, our interpretations are therefore inherently subjective (indeed a potential hazard, but one that proper methods and good researchers account for).

While some of these objections are true (and are why we don't recommend reporting numbers from qualitative studies), it's a big jump to assert that qualitative research is anecdotal or lacks rigor. It is simply the case that [qualitative research is a rather different mode](#) of investigation.

If you are a UXer that is facing this sort of pushback, consider making the following points to your coworkers.

## Qualitative Methods Are a Necessary Complement to Quantitative Measurement

People with a science education are usually familiar with experiments that use [carefully controlled quantitative measurement](#) as a way of evaluating hypotheses; this is typically known as *probative* research. For people with this background, the idea of talking to a small number of people (and perhaps even changing the study procedure slightly each time!) as a means of drawing conclusions may seem inherently unscientific, prone to bias, and unlikely to generalize to the general population.

But the goal of qualitative research is different: we're not trying to disprove a hypothesis, but we're looking to understand the nature of a problem in detail. [Qualitative](#) research doesn't try to make quantitative claims that generalize to the whole target audience. Just because 6 people in a 10-person study are able to, for example, easily use a feature in an app does not mean that we can say that 60% of the overall population will have a similar experience. But, in that study, we can identify the issues that the 4 other people encountered (and also those that the 6 others struggled with, yet overcome) and understand the reasons behind

them, with the goal of fixing those problems. (These issues are something that [we can only speculate](#) about if we look solely at a [quantitative study or analytics data...](#))

Different goals require [different methods of investigation](#): knowing how many people are having a problem requires a large sample size to be confident that the number we measure isn't [distorted by random chance](#), but knowing **that a problem may occur** and **why** requires that we observe behaviors and elicit users' thoughts. Most important, **how** to redesign the UI to fix the problem requires these qualitative insights.

The goal of qualitative research is to gather insights that drive decision making, especially when measurement is unfeasible or impossible. While, of course, we can come up with some type of survey instrument for measuring [satisfaction](#), [emotional state](#), and other [internal phenomena](#), those tools don't tell us *why* the user feels that way at the moment or *how we can better support their needs*.

## Qualitative Research Is Rigorous and Systematic

Still, an important question is: how can we know if qualitative research is rigorous and dependable to give us true insights about our users?

Rigor in quantitative research is seen as being comprised of a few major attributes:

1. [Validity](#) — is the thing we're measuring a good representation of the thing we care about? Can our conclusions generalize beyond this experiment?
2. Reliability — if we repeat the research, will we get similar results?
3. Objectivity — do we have a way of ensuring that our observations aren't clouded by our biases?

These characteristics are relatively straightforward for quantitative research, but are not easy to establish for most studies with small sample- sizes.

Social scientists Yvonna Lincoln and Egon Guba created a parallel set of characteristics for qualitative research that have become a standard way of

## assessing rigor:

1. Credibility: Did we accurately describe what we observed?
2. Transferability: Are our conclusions applicable in other contexts?
3. Dependability: Are our findings consistent and repeatable?
4. Confirmability: Did we avoid bias in our analysis?

We can satisfy those criteria by being systematic. That is the factor that makes the data we collect *data*, not anecdotes that happen by chance. If the CEO hears from a friend that the company's app looks outdated, that's an anecdote — there wasn't a systematic process to gather that observation, it happened by chance and is only one person's subjective opinion. If a UX researcher systematically recruits 5 participants and several of them struggle to understand the branded terms in the navigation, that is data.

Good qualitative researchers take many steps to ensure their work is systematic:

- They relate their work to an evidence-based theoretical [framework](#) about how UIs should be designed and to a cognitive-psychology and human-computer–interaction body of knowledge about how users [sense the world and mentally process it](#), [behave](#), and engage in [specific interactions](#) with various forms of technology.
- They formulate [specific research questions](#) before choosing the appropriate method.
- They carefully sample by [recruiting participants](#) that represent a variety of perspectives, so they can learn about unknown unknowns.
- They facilitate sessions using [open-ended prompts](#) to elicit participants' thoughts and reactions with minimal bias and following up on intriguing incomplete statements from participants without [unconsciously indicating to them how to respond](#).
- They don't just take users' [opinions at face value](#) — they build up an understanding of *why* a user may request a feature or *why* something might not look appealing, for example.
- They analyze our data by [systematically coding the insights](#) (again, drawing on the overarching theoretical framework of heuristics, [known best practices](#), and so on). Coding is often done using inductive reasoning techniques

borrowed from grounded-theory methodology — themes emerge from the data in a bottom-up analysis, rather than by starting with a list of codes that is then force-fit to the data. They then try to establish conceptual connections between the coded findings in the data [and look for patterns](#).

- When they encounter something unusual or extraordinary, they use [triangulation](#) to ensure that our conclusions are supported (i.e., they investigate the same thing via a different method or they have other trained researchers analyze the same data independently). Extraordinary claims require extraordinary evidence, after all.

## Small Sample Sizes Are Fine, Depending on What You're Looking At

But, you may be saying, [What about those small sample sizes?](#) Don't they have an inherent [sensitivity to outliers](#)? Maybe a problem you observe is real, but rare, and you might [overstate its importance due to a small sample](#).

These are all real concerns. So, how *do* qualitative researchers protect against that sort of overrepresentation of rare events in our conclusions?

Once again, we can point back to a robust [theoretical framework](#) we have in UX full of evidence-based principles about how users sense, think, behave, and interact with technology. If we observe even one person having a problem that is an exemplar of a **known** principle, we are able to be reasonably confident that it is a real problem. Of course, we still won't be able to say precisely how many people will encounter that problem.

If the number of people affected by a problem is a real factor that we need to consider (e.g., if the problem will be expensive to fix and will take a lot of resources), then yes, we may need to do some form of quantitative experiment to figure that out. On the other hand, it is often cheaper (and more sensible) to simply fix the design problem without quantifying just how bad it is, *if* we've identified it [early in the design process](#).

For example, if I design a deep fryer meant for consumer use, I will (hopefully) do

some safety testing before selling it. If the first tester accidentally burns themselves on the fry basket because the handle is right over a heating element, I will probably not keep testing it with a large sample to figure out exactly what proportion of users will also burn themselves and sue me. In this case, I've found a major problem with a sample size of 1. Now, this example is obviously very simplified, and a small sample size will not be appropriate for every research question, but this approach will be often the [best use of resources](#), especially when we're looking for major blockers.

That is one of the main reasons we have consistently [recommended small sample size studies](#), done early (and [repeated with several iterations of a design](#)): they are a relatively [inexpensive](#) way to find and address major usability issues that we would otherwise learn about from angry customers if we shipped the product without testing. It would be a waste of time and resources to confirm a major flaw in the design with many participants, especially if we're working on a fast-moving Agile team.

A reasonable question might be asked: Why not use simply larger sample sizes for qualitative studies, in order to be more confident that, for example, the expectations and needs expressed by our study participants are commonplace, not unusual outliers? Fundamentally, this comes down to cost: to recruit more participants and also to moderate the study sessions. Eliciting users' inner thoughts often requires a [skilled facilitator](#) who may need to do some improvisation during each session in order to adjust to the specifics of each study participant. Moreover, because the session protocol will be a little different every time, we would not be able to soundly compare and aggregate the data from all the sessions, since each "trial" would be different. In practice, qualitative researchers often land on a specific sample size based on how many participants it takes to reach a saturation point in the findings (i.e., they continue the study in small batches until they're unlikely to learn enough new core insights to be worth the added delay to the project). Especially for [interviews](#), [field studies](#), and other forms of [discovery-oriented research](#), this is the goal — rather than trying to determine how *common* the core findings are.

# Empathy and Humanity Aren't Easily Counted, but They Count

Last, but definitely not least, qualitative research allows us to build a real, [empathetic understanding](#) of users as human beings. When we view human interactions with technology primarily through the lens of metrics such as engagement, bounce rate, or time on task, we aren't very concerned with the users' well-being. (It might be in the back of our minds, but certainly not a primary consideration.) The tech industry is just beginning to reckon with the [ethics](#) of what we do and to realize that how we design our products has a real impact on the life of many, many human beings.

[Moderated](#) qualitative research requires that we engage with other humans (and even [unmoderated](#) studies still involve observing people). We typically need to build some form of [rapport](#) to get participants comfortable with expressing their inner thought process. We often discover that they experience the world differently than we do — in ways both small and subtle, and huge and overt. These studies provide the *opportunity* to empathize with them.

I don't want to overstate the power of qualitative research here. It will not automatically generate empathy for users — I've certainly witnessed teams laughing while watching users struggle. Doing qualitative research will not fix ethical problems baked into a business model. Qualitative research certainly will not replace the critical need for just and inclusive hiring practices for your team, to ensure that decisions are made by people with a variety of backgrounds and lived experiences.

On the other hand, I also don't want to *undersell* the value of the empathy built through this sort of research — for example, simply through noticing how frustrated one user gets and hearing them [casually questioning if they are stupid](#) because they couldn't figure out a confusing design. That (unfortunately commonplace) reaction tells me that the problem is real and fixing it needs to be a priority, even if I don't have a huge sample size.

## Summary

Qualitative research is rigorous and systematic, but it has different goals than quantitative measurement. It illuminates a problem space with data about human experience — expectations, mental models, pain points, confusions, needs, goals, and preferences. Sample sizes are typically smaller than for quantitative experiments, because the goal isn't to suggest that our sample participants will represent the whole populations proportionally; instead, we're looking to find problems, identify needs, and improve designs. UX research is a mixed-methods discipline because these two approaches are complementary: measuring *how much* and understanding *why* can both help us build better products, which is the main goal of any UX research.

## References

- Juliet Corbin and Anselm Strauss. 1990. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*, Vol. 13, No. 1, 1990
- Yvonna Lincoln and Egon Guba. 1985. *Naturalistic inquiry*. Sage, Newbury Park, CA.
- Saunders, B., Sim, J., Kingstone, T. *et al.* 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 52, 1893–1907. <https://doi.org/10.1007/s11135-017-0574-8>
- Mike Hughes. 2011. Reliability and Dependability in Usability Testing, Retrieved from: <https://www.uxmatters.com/mt/archives/2011/06/reliability-and-dependability-in-usability-testing.php>



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