

Quantitative versus qualitative user research.

Part 1: Swings and roundabouts.

Maxwell J. Roberts
July 2024

Downloaded from:
www.tubemapcentral.com/writing/webarticles/Roberts_QversusQmethods_2024.pdf

Also at LinkedIn:

<https://www.linkedin.com/pulse/quantitative-versus-qualitative-user-research-part-1-swings-roberts-kmgce/>

More articles at:

www.tubemapcentral.com/writing/writing.html

Many people conducting user research suggest that quantitative research methods have certain issues and that these can be resolved by using qualitative methods instead, with the added bonus of richer data and deeper insights into human behaviour. In this article, I suggest that qualitative research does not solve the alleged issues of quantitative research, instead these are merely camouflaged. Furthermore, collecting richer data comes at a cost. The vast, unstructured datasets that result from qualitative research comprise exactly the sort of information that humans find particularly difficult to interpret.

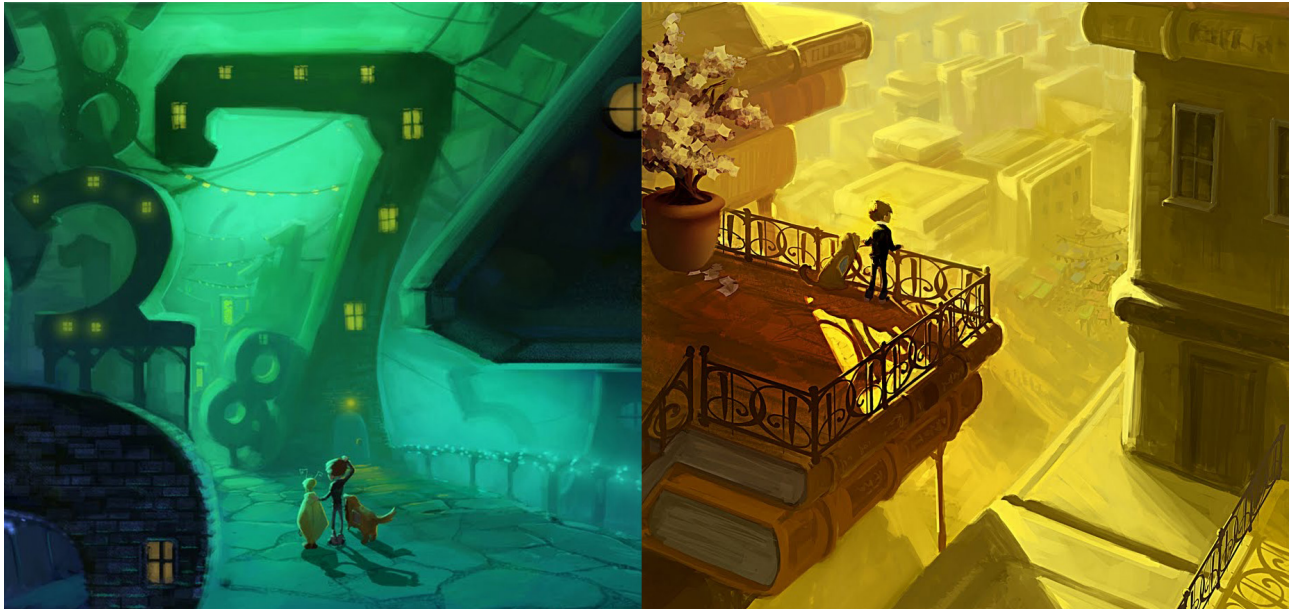


*Images depicting scenes from The Phantom Tollbooth created by Lizzie Nichols, 2010,
<http://lizzienichols.blogspot.com>.*

In behaviour research, there is schism between quantitative versus qualitative methods, with many people leaning towards the latter. Briefly:

Quantitative methods involve converting human behaviour and beliefs into numbers, such as the time taken to perform a task, numbers of choices for different options, numbers of errors, numbers of different types of error, disposition and belief levels (for example, by scoring statement ratings tasks) and so on. This methodology is most powerful when applied to A/B testing: which version/prototype is easier to use, which version is liked more? No interface is perfect, and people will always make mistakes, but we can find out which version is 'better' according to many different dimensions. If we devise our tasks well enough and still can't find any differences, then the chances are that, in the real world, it won't make much difference which prototype is adopted.

A good example of quantitative methodology applied to the real world is research that I led evaluating prototype Docklands Light Railway route diagrams. Here, a map can be considered to be an interface between a user and the transport network. We looked at objective measures of performance (for example, time to find named stations, time to plan journeys between pairs of stations). We also asked people to identify preferred maps and to rate these on numerical scales according to various criteria. Hence, this can effectively be considered as an example of UX research. The article in which this research is published can be downloaded here: http://www.tubemapcentral.com/writing/academicfiles/Roberts_Rose_TRA_2016.pdf



In the classic children's story, The Phantom Tollbooth by Norton Juster, the unlikely hero, Milo, visits Digitopolis, the kingdom of numbers (left) and Dictionopolis, the kingdom of words (right). The kingdoms are at war and Milo must reconcile their differences. Quantitative and qualitative researchers are not at war, but there are definitely disagreements between the two camps about the efficacy of these methods. These splendid images were created by Lizzie Nichols, 2010.

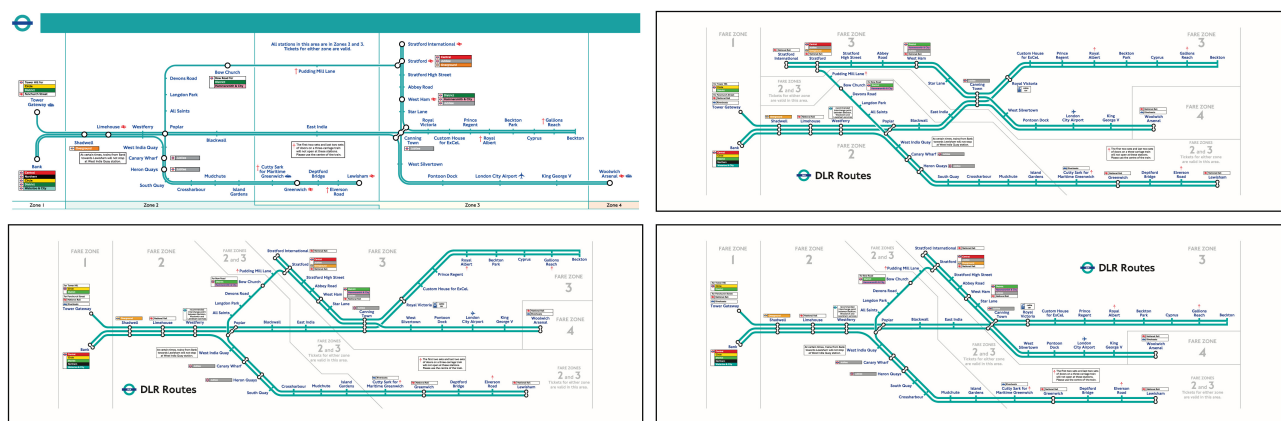
Qualitative methods involve collecting lots of words from people while they are performing various tasks. For example, in a usability testing context, asking people to verbalise what they are trying to do, how they are going about doing it, and their expectations and feelings along the way. These will be supplemented with observations from testers. The key defining feature of this methodology is that no attempt is made to quantify what happened, not even how often something went wrong, because then the research would become quantitative. Numbers are forbidden. The argument is that this raw, unimpeded window on the soul gives us richer and deeper insights into people's interactions and dispositions.

I am experienced with using quantitative methods to investigate usability, having designed and conducting numerous studies that have yielded all sorts of (often subtle) insights into people's performance. These also enable the identification of differences between people in what information they need and the strategies they prefer to apply. Taking into account the diversity and ingenuity of quantitative methodologies devised over the years by experimental psychologists, I often wonder whether some of the claimed advantages of qualitative methods might be overstated. "Experience at qualitative methods" is the mantra for so many UX design job advertisements, but do the people writing this sort of thing really understand the advantages and disadvantages of what they are asking for?

Many people strongly advocate the use of qualitative methods and the motivation for this is often that there are alleged shortcomings of quantitative methods. However, I am going to argue that, for the two issues most often identified, qualitative methods are not necessarily the answer: instead these merely camouflage the shortcomings, giving a false impression that these have been adequately addressed. Qualitative methods can even add their own problems to the mixture.

1) Quantitative methods are data hungry. If qualitative methods are used instead then their richer data has the consequence that fewer people need to be tested to get the same, or better insights.

When we applied quantitative methods to evaluating Docklands Light Railway map prototypes, this absolutely enabled us to identify the best of the prototypes. However, to reach these (statistically significant) findings we needed to test around 280 volunteers, each for 30 minutes. Could we have identified the best prototype using a qualitative study with fewer people? Only if we were very lucky, as we will see.



The four prototype Docklands Light Railway maps tested on behalf of Transport for London. The winning design, with the most efficient planned journeys and the fewest errors, is top right.

The maps we tested were *basically competent*: created by designers who knew what they were doing. This meant that errors made by users when navigating with them were relatively rare. Even so, we wanted to minimise any bad decision making, by users, that might be induced by these maps. However, user error is inevitable no matter how carefully anything is designed, so it is unrealistic to aim for a zero-error design, only a lower-error design in relation to the other prototypes tested. Our objective, therefore, would be to find the best design by *counting the number of errors for each one*. By definition, it is hard, therefore, to imagine how we could even address this objective with a qualitative study, but let's try. We had four map prototypes to test in our study, so let's give one map to each of five people, and we will ask them to talk us through planning some journeys. Unfortunately, given that errors are relatively rare — say the chance that someone will make a mistake using the most error-prone map is 10% — then if we test five people on this map, roughly 60% of the time everyone will perform perfectly and the design defect(s) will go unnoticed.

More generally, moving into familiar UX research territory, provided an interface is not inherently utterly defective, most people will perform reasonably well most of the time. Each potential trouble-spot is unlikely to catch out many people, and perhaps no one at all if your study is too small. The problem, therefore, is not that quantitative methods, in particular, Hoover up a lot of people. Instead, the problem is that unless the interface under evaluation is disastrously poor, you are going to need to test a lot of people in order to have a good chance of catching as many of the trouble-spots as possible. Subtle defects require large-scale testing to detect them, no matter whether you choose qualitative or quantitative methodology. Conversely, if your interface has such glaring design flaws that most people trip up when attempting to use it, then you won't need a large-scale quantitative study to detect this either, a small-scale one will suffice. Overall, the problem is not that quantitative methods are data hungry, it is that *subtle defects are data hungry*.

2) Quantitative studies are reductionistic, shallow and unsubtle. *If qualitative methods are used instead, their richer data will yield a better and deeper understanding of usability issues.*

This argument is, essentially, that compressing human behaviour down to numbers squashes the insights out of it, resulting in potentially important issues being missed. There is some merit to this argument: performance at some tasks can be difficult to quantify and, for complex multi-stage tasks, it might be the case that a qualitative talk-me-through-everything usability study is all that is feasible. However, experimental psychologists are expert at using all sorts of methodological tricks to coax subtle quantifiable behaviour from volunteers that should be able to answer key usability questions in many instances.

Qualitative data is undoubtedly rich, but the resulting stack of transcriptions of verbal statements made by volunteers while solving tasks, comprising whatever was going through their heads at the time, alongside additional observations by the testers, is vast and unstructured. The statements might be contradictory between people if there are individual differences in ability and preferences. The statements might even be self-contradictory if, for example, the interface is complex to use. How can researchers make sense of these in order to identify, for example, which of the prototypes is the most effective, and which issues on the preferred version need to be fixed, and how?

I have over three bookshelves devoted to the topic of how people make sense of information and draw conclusions from this. The research more or less unanimously implicates that the data from qualitative research is exactly the sort of complex, sprawling, multidimensional information that will be cognitively overwhelming. Even experts will find these data to be difficult to interpret and identify valid conclusions from. Faced with this complexity, they will be susceptible to all sorts of biases and systematic errors of inference and judgement. The most well-known of these is confirmation bias, in which people with strong preconceptions seek out, selectively attend to, and highlight information to others that validates their prejudices. A vast disorderly dataset provides ample opportunity for this.



My bookshelves on the psychology of inference and information interpretation. The cognitive load presented by sprawling, disorganised information will be immense, even for experts at using qualitative methodology.

Quantitative research and statistical testing were not invented because researchers wanted to oversimplify the world with impoverished data. Instead these originated from a desire to improve our ability to make sense of the world by implementing an orderly framework within which tools would be available to help make

sense of observations, reducing demands on limited cognitive resources and reliance on potentially fallible, subjective, biased judgements.

Presented like this, there is a clear dissociation between qualitative and quantitative research: quantitative research yields relatively simple, orderly data that is reasonably straightforward to make sense of, whereas qualitative research yields relatively complex, disorderly data that is much harder to make sense of, which will inevitably impact the sophistication of the conclusions that can be drawn from it. You might (perhaps cynically) summarise this as:

Qualitative research: simple conclusions from rich data

Quantitative research: rich conclusions from simple data

If you accept this analysis, then it follows that quantitative research should always be the first choice unless something about the tasks being investigated points towards the need for some sort of qualitative approach. For example, a quick-and-dirty analysis might be needed of a single prototype to make sure that there are no serious usability issues, or else research might be desired that ventures into new ground and researchers are not sure what to expect. Another way of characterising the difference between qualitative and quantitative methods, and emphasising that both have a role to play, is that qualitative studies yield hypotheses, quantitative studies yield evidence.

In part two of this article I will take a look at an analysis methodology that is often used in an attempt to make sense of qualitative data: affinity diagrams.