

User Testing

By Richard Neves

WHAT IS IT?

User testing is the gold standard empirical approach that other methods are compared to. The method collects data about the user's performance with a design/product. It involves observation and direct query. User testing is key to User Centred Design (UCD).

WHY DO YOU NEED IT?

As Nielsen said in 2001, *'pay attention to what users do, not what they say'*. There are many reasons why user testing is important:

- Because we cannot accurately predict how users will find the product.
- To learn whether it meets the user needs.
- Designers are not users, have they accurately complemented the user mental model.
- Ensures issues can be fixed before going to market.

WHEN DO YOU NEED IT?

User testing can happen at any stage in the product life cycle, formative (feeds back into the design throughout the development cycle, identifying usability problems) and formative (at the end of the process, metrics and benchmarks).

HOW DO YOU DO IT?

There are several stages ahead of the testing. Dumas and Redish (1999) came up with a 10-step framework:

1. **Define goals and concerns driving testing** – the testing needs to have a clear purpose or you won't know if you have an answer or way forward. You may have qualitative testing to answer whether users can navigate effectively, or a declarative statement around quantitative testing, e.g. 90% of users will be able to save their search query without any mistakes. The origins of the testing goals may have come about through many sources, including: internal feedback, expert reviews or previous usability tests.

2. **Decide who should participate** – clearly testing your design on anyone is better than no one, but in reality you want participants who are your intended users. In some cases, for commercial websites for example, this can be largely down to demographics, but you should also consider domain knowledge, context of use, experience of similar, computer competency, motivation for use and physical ability.
3. **Recruit participants** – you need to think about the inclusion/exclusion criteria, what are the distinctions that make users behave differently, i.e. content knowledge rather than just demographics. But don't only think about current users, consider future users too. Recruitment could be through client or personal contacts, sales and marketing teams, opportunistic (guerrilla/corridor testing) or agency recruitment. However, ensure you have prepared a screener questionnaire so that you are recruiting participants that are current or potential users. The screener may include questions on length of time in a specific role (are you looking for novices or experts?), or to find out more about their role and a confirmation that the individual works in the industry of interest to you. You can also ask about what an individual does on their computer/smartphone to get an idea of how computer literate they are. Essentially depending on who you need to meet your study objectives, tailor your screener to maximise the potential for testing the right type of user.

If you are recruiting by phone or email, be sure to prepare a script or email to tell the potential participant what you are doing, why you are doing it, what is asked of the participant, what data is collected, that the system being tested not them and any incentives. When the participant accepts,

send a follow-up confirmation email with details of session time, location etc. and also attach the consent form, it's good to have this signed ahead of the session, but if not, at least the participant is aware of it and can sign at the beginning of the session.

Don't forget to send a reminder email a week before the session.

How Many Participants?

This is the question that's always being asked (or ignored) and while there are no absolutes, here's a simple guide that I tend to stick to:

	Benchmark	Formative	Summative
Participants	8-24	4-6	6-12
Metrics	Time, error	Qualitative	Metrics based in goals
Why	Baseline	Find/fix issues	Measure success
When	Early – before designing	During design	End of the design process
Frequency	Once	Multiple	Once

Reddish and Dumas (1999)

There are several factors involved in determining the number of users, but I find the following two aspects key:

- *Novelty of the System* – something that is a first will probably require more research and therefore time as there is a good chance of more design iterations.
- *Criticality* – what is it you are testing? The extent of the research is likely to vary if you are testing a control panel for hospital use versus a museum information kiosk.

4. **Select and organise tasks to test** – If the tasks are poor, then your testing may not successfully meet its objectives. Ensure the tasks test the relevant parts of the design and areas of concern. Develop tasks that are realistic and meaningful to the participant, remember that for some roles the detail is important, so be sure to get the tasks accurate to avoid any problems during the testing.
5. **Create task scenarios** – you need to think of a typical, short scenario to set the context and express the goal from the participant's perspective, using language that resonates with them. Do not provide the steps in the task! Avoid

scavenger hunt tasks (Spool, 2006); these are simplistic tasks asking participants to find something on the site. Everyone will eventually find it and it doesn't tell you very much. When writing your tasks, consider the following:

- Start with something not too challenging, to ease the participant into the session.
- Provide the tasks in logical order.
- Cover the important tasks early on.
- Ensure all tasks are achievable within 45 minutes (avoid testing sessions lasting more than an hour – fatigue will set in and the participant may not want to help out again in the future).

6. **Decide how to measure usability** – remember that qualitative testing generally leads to some redesign: problems observed, opinion given. Quantitative testing will give you values, but doesn't really tell you what to change. You can collect all sorts of data:

- interaction with the system, e.g. observation and Morae interaction recordings
- Verbal behaviour, i.e. thinkaloud
- Non-verbal/expressions, e.g. video recording
- Opinion, e.g. questionnaires
- Eye movement, e.g. Tobii eye tracking
- Physiological measures, e.g. galvanic skin response (skin conductivity) and ECG (heart rate monitoring).

Most commonly we record interaction with the system, verbal behaviour and opinion.

7. **Prepare other materials for test** – it's important that you have something ready to test, the fidelity depends on what your objectives are:

- **Low fidelity** – e.g. sketches, paper prototypes, wireframes. Although these are cheap and quick to prepare, they are not so good for testing interactivity.
- **High-fidelity** – e.g. Axure or html. These are more time consuming to prepare, but are great for assessing complex dynamics and functionality.
- **Running system** – of course you can test on a running system too, but be sure you are aware of any limitations and ensure it's running for the testing session.

8. **Prepare test environment** – this covers everything from booking the testing room and hiring equipment to printing out materials for the session:
 - Screener questionnaire
 - Participant invitation, confirmation and reminder email
 - Welcome information sheet – about the session and the order of the activity, what they'll do, the data collected, system being tested not them etc.
 - Consent form
 - Test script and any task scenario cards for the participant, for example if you want them to type something specific into a form etc.
 - Pre-test questionnaire
 - Training material if necessary (not normally)
 - Prototype or running system
 - Post-test questionnaire
 - Data recording forms for you and the observer to take notes
9. **Prepare test team** – ensure that if you have observers they understand the flow of the session, how and what to record. If technology is being used ensure the team know how to use it.
10. **Conduct a pilot test** – it is always recommended that before you run a testing session with a participant that you have gone through the test in real time. This is important because:
 - It ensures that the set-up and materials are prepared and fit for purpose
 - Creates a practice run for the team
 - Ensures that testing satisfies goals/concerns
 - Also ensures that you can run the test in the allotted session time.
3. Administer the pre-test questionnaire – this essentially confirms the user being tested is the right one.
4. Provide training of the system if necessary; also explain thinkaloud and how you may probe the participant if they go quiet. Also depending on the kind of test you are running, you may need to explain that you cannot answer any questions during the session, but can address these at the end of the session. Ahead of the testing, ask the participant if they have any questions before proceeding.
5. If you're recording the session, begin this now.
6. Begin the testing and give the participant their first task.
 - Remind them to thinkaloud
 - Observe and take notes
 - If the participant is stuck, allow them to explore a bit more, but not for too long or they'll become frustrated and become less motivated for the rest of the testing. Move to the next task, tactfully.
7. When all the tasks have been given, conduct the debriefing interview and respond to any questions asked that you couldn't answer during the session.
8. Administer the post-questionnaire.
9. Thank the participant and reward them with their incentive if you have offered one.
10. Now meet with your team and consolidate notes and summarise findings as soon as possible.

Writing up the testing results

For all the usability problems identified:

- Describe each problem and what caused it. But first decided what is a usability problem? When a user fails the task? User articulates the goal but doesn't achieve it until two or more tries? Participant takes the suboptimal path? Participant achieves success unexpectedly?
- Group and categorise like related issues
- Rank the severity of the problem, 1-4 (cosmetic to critical)
- Finally, make design recommendations based on design principles or patterns, e.g. does the feature break design guidelines/familiar pattern standards? Is there a mismatch

THE USABILITY TEST

The use of a script will always ensure that you provide the same instructions to all participants, a kind of controlled session.

1. Welcome participant – it's important to put the participant at ease, provide them with the session and practical information, e.g. that you are testing the system and not the user and where the facilities are located (as well as the exit), respectively. Use your welcome information sheet to ensure you don't forget anything.
2. Obtain informed consent.

between the design and the user's mental model? Is there a cognitive load or learning issue?

The structure of the testing report depends on your organisation or client. In many cases it will be a slide deck presenting the highlights. A more formal report will usually take the following structure:

- Title page
- Executive summary
- Introduction – product description, test objectives
- Method – the participants, context, tasks, facility, equipment and study design
- Usability metrics – for example, effectiveness, efficiency and satisfaction
- Results – data analysis, tables, performance data
- Appendices – letters to participants, instructions, tasks, questionnaires

WANT TO KNOW MORE?

More detail on User testing:

A Practical Guide to Usability Testing

Dumas and Redish

Don't Make Me Think

Steve Krug