Jakob Nielsen's Alertbox, March 19, 2000:

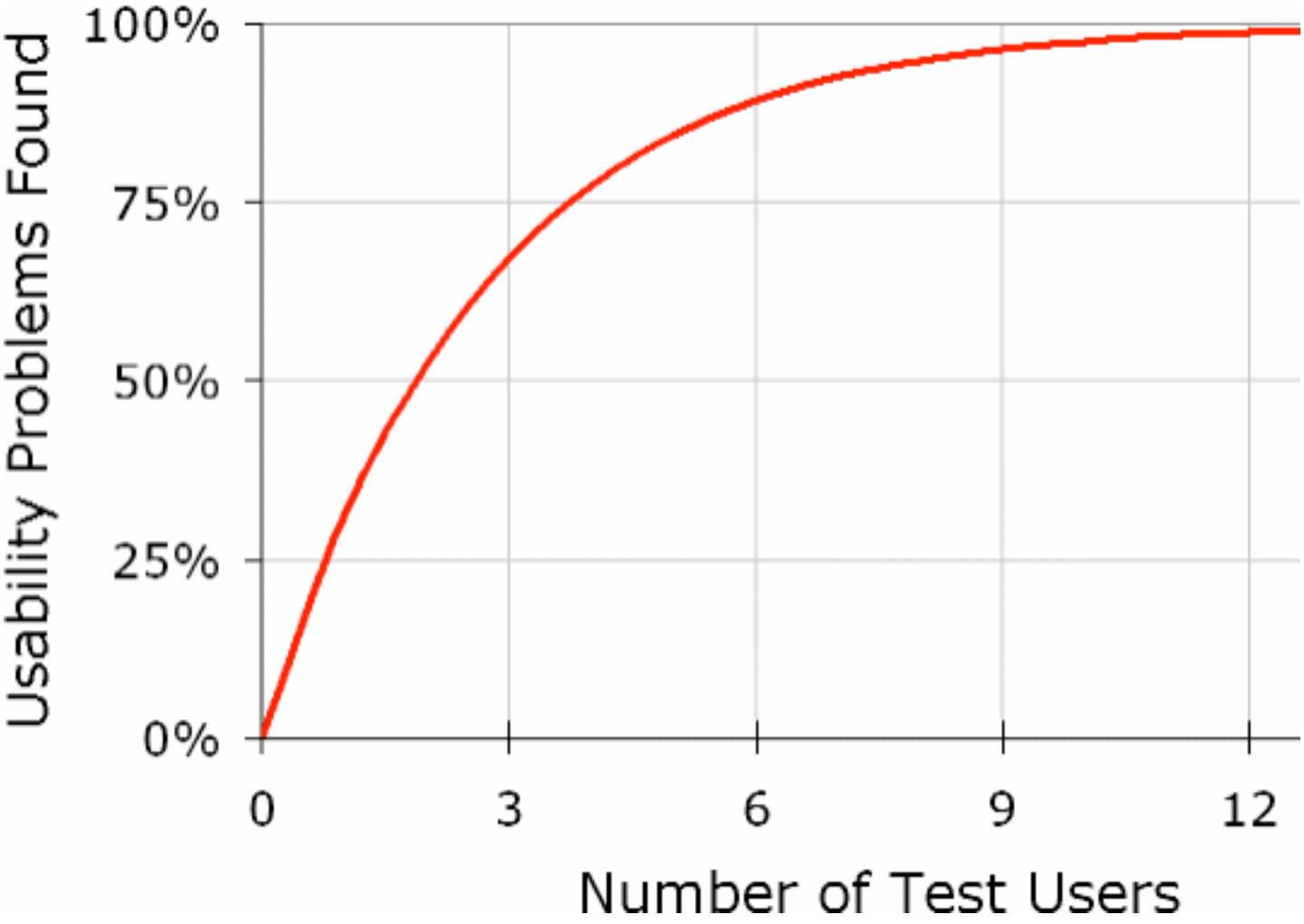
Why You Only Need to Test with 5 Users

Some people think that usability is very costly and complex and that user tests should be reserved for the rare web design project with a huge budget and a lavish time schedule. Not true. Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford.

In earlier research, Tom Landauer and I showed that the number of usability problems found in a usability test with *n* users is:

$$N(1-(1-L)^n)$$

where *N* is the total number of usability problems in the design and *L* is the proportion of usability problems discovered while testing a single user. The typical value of *L* is 31%, averaged across a large number of projects we studied. Plotting the curve for *L*=31% gives the following result:



The most striking truth of the curve is that **zero users give zero insights**.

As soon as you collect data from a **single test user**, your insights shoot up and you have already learned almost a third of all there is to know about the usability of the design. The difference between zero and even a little bit of data is astounding.

When you test the **second user**, you will discover that this person does some of the same things as the first user, so there is some overlap in what you learn. People are definitely different, so there will also be something new that the second user does that you did not observe with the first user. So the second user adds some amount of new insight, but not nearly as much as the first user did.

The **third user** will do many things that you already observed with the first user or with the second user and even some things that you have already seen twice. Plus, of course, the third user will generate a small

amount of new data, even if not as much as the first and the second user did.

As you **add more and more users, you learn less and less** because you will keep seeing the same things again and again. There is no real need to keep observing the same thing multiple times, and you will be very motivated to go back to the drawing board and redesign the site to eliminate the usability problems.

After the fifth user, you are wasting your time by observing the same findings repeatedly but not learning much new.

Iterative Design

The curve clearly shows that you need to **test with at least 15 users to discover all the usability problems** in the design. So why do I recommend testing with a much smaller number of users?

The main reason is that it is better to distribute your budget for user testing across many small tests instead of blowing everything on a single, elaborate study. Let us say that you do have the funding to recruit 15 representative customers and have them test your design. Great. **Spend this budget on three tests with 5 users each!**

You want to run multiple tests because the real goal of usability engineering is to improve the design and not just to document its weaknesses. After the first study with 5 users has found 85% of the usability problems, you will want to fix these problems in a redesign.

After creating the new design, you need to **test again**. Even though I said that the redesign should "fix" the problems found in the first study, the truth is that you *think* that the new design overcomes the problems. But since nobody can design the perfect user interface, there is no guarantee that the new design does in fact fix the problems. A second test will discover whether the fixes worked or whether they didn't. Also, in introducing a new design, there is always the risk of introducing a new usability problem, even if the old one did get fixed.

Also, the second test with 5 users will discover most of the remaining 15% of the original usability problems that were not found in the first test. (There will still be 2% of the original problems left — they will have to wait until the third test to be identified.)

Finally, the second test will be able to **probe deeper into the usability of the fundamental structure** of the site, assessing issues like information architecture, task flow, and match with user needs. These important issues are often obscured in initial studies where the users are stumped by stupid surface-level usability problems that prevent them from really digging into the site.

So the second test will both serve as quality assurance of the outcome of the first study and help provide deeper insights as well. The second test will always lead to a new (but smaller) list of usability problems to fix in a redesign. And the same insight applies to this redesign: not all the fixes will work; some deeper issues will be uncovered after cleaning up the interface. Thus, a third test is needed as well.

The ultimate user experience is improved much more by three tests with 5 users than by a single test with 15 users.

Why Not Test With a Single User?

You might think that fifteen tests with a single user would be even better than three tests with 5 users. The curve does show that we learn much more from the first user than from any subsequent users, so why keep going? Two reasons:

- There is always a risk of being misled by the spurious behavior of a single person who may perform certain actions by accident or in an unrepresentative manner. Even three users are enough to get an idea of the diversity in user behavior and insight into what's unique and what can be generalized.
- The cost-benefit analysis of user testing provides the optimal ratio around three or five users, depending on the style of testing. There is always a fixed initial cost associated with planning and running a test: it is better to depreciate this start-up cost across the findings from multiple users.

When To Test More Users

You need to test additional users when a website has **several highly distinct groups of users**. The formula only holds for comparable users who will be using the site in fairly similar ways.

If, for example, you have a site that will be used by both children and parents, then the two groups of users will have sufficiently different behavior that it becomes necessary to test with people from both groups. The same would be true for a system aimed at connecting purchasing agents with sales staff.

Even when the groups of users are very different, there will still be great similarities between the observations from the two groups. All the users are human, after all. Also, many of the usability problems are related to the fundamental way people interact with the Web and the influence from other sites on user behavior.

In testing multiple groups of disparate users, you don't need to include as many members of each group as you would in a single test of a single group of users. The overlap between observations will ensure a better outcome from testing a smaller number of people in each group. I recommend:

- 3-4 users from each category if testing two groups of users
- 3 users from each category if testing three or more groups of users (you always want at least 3 users to ensure that you have covered the diversity of behavior within the group)

Reference

Nielsen, Jakob, and Landauer, Thomas K.: "A mathematical model of the finding of usability problems," *Proceedings of ACM INTERCHI'93 Conference* (Amsterdam, The Netherlands, 24-29 April 1993), pp. 206-213.

Read More

- Quantitative Studies (usability metrics): Test 20 Users
- Card Sorting: Test 15 Users
- 230 tips for usability testing
- 233 tips for recruiting test users
- Return on Investment from Usability
