



# HFI User Experience Design Newsletter

New Techniques: How to Talk Truthfully About Usability Testing

Message from the CEO, Dr. Eric Schaffer

May, 2010

[Newsletter](#)



**Human Factors**  
International

# HFI User Experience Design Newsletter

May, 2010

---

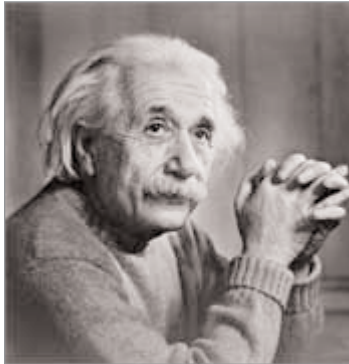
## Table of Contents

---

<b>New Techniques: How to Talk Truthfully About Usability Testing</b> by John Sorflaten, PhD, CPE, CUA, HFI Writer	<b>3</b>
<b>Message from the CEO, Dr. Eric Schaffer</b> by Eric Schaffer, PhD, CPE, CUA, Founder and CEO	<b>14</b>

---

## New Techniques: How to Talk Truthfully About Usability Testing



Is it possible to simply "be too smart"? I paged through a book once on how to teach genius kids - got it at a garage sale.

The most disappointing phrase I found was that "there's a limit to how smart a leader can be."

What? Explain THAT please!

Well, to make you all feel better, here's a nice quote from an award-winning researcher, Dean Keith Simonton.

*The relationship between IQ and persuasive influence over other members of one's group may drop off beyond an IQ of approximately 120. Individuals who are very smart may be less comprehensible to other group members, and this might have a negative impact on their ability to influence those around them....*

*As evidence, Simonton's historiometric research demonstrates that more intelligent individuals have had a harder time getting elected to the American presidency! (Simonton, 1999a)*

### Are you just too smart to sell usability?

No doubt, you have been wondering why it's been a challenge to sell usability services to your colleagues. Well, now you have an answer. Maybe you are simply too smart.

Remember, if you have an IQ over 120, you have a communication problem.

Now that I have your attention, you can appreciate hearing about major advances that have recently occurred in explaining usability testing results.

We have people like usability statisticians James Lewis and Jeff Sauro to thank for their breakthrough efforts in creating simple, down-to-earth phrases for explaining your usability testing results.

Even better, their quest for simplicity offers a better grasp of the truth behind your usability testing results.

Now you have tools to clearly explain your usability test results. Are you tired of saying "we found 85% of the problems with 5 users." (In fact, that's probably wrong.)

The new "talk the talk" lets you mitigate risk, quantify your progress, and justify your work in usability.

Since you have an IQ over 120, I have no fear about using words like "mitigate" or "quantify" or "usability".

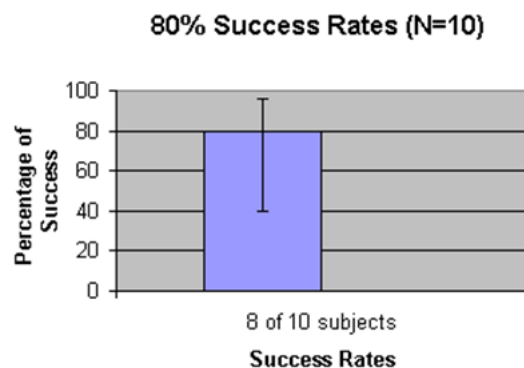
Let's see how Jeff Sauro and James Lewis have opened the door to down-to-earth talk about truth in testing.

### The old way of talking about usability testing

I get to complain about my own former attempts at explaining usability testing results. (Yes, I am that objective J ) I flunked in my mission.

Check out my article When Discount Usability Misleads Management - A Solution originally published in the HFI UI Design Newsletter on Jun/July, 2005. I advocated that you should clearly explain that your "pass rates" on testing include the "margin of error".

Here's the classic picture, for which you would notice that a pass rate of 8 out of 10 participants must include mention of that "95% confidence interval". This means that if you repeated the test protocol 100 times, then the pass rate will fall between the bottom and the top margins of error in 95% of those occasions.



I even showed you a neat calculator recently created by Jeff Sauro on his great web site [www.measuringusability.com](http://www.measuringusability.com). It calculated a special version of the confidence interval that handled small numbers of subjects. Here's what I told you:

**Use the Adjusted Wald results because they take into account the small number of subjects (under 150).** Note the lower bound is now 48.29% (48%) instead of 55%. The upper bound is now 95.62% (96%) instead of 105%.

### Yikes, I flunked the IQ test!

Clearly, in my enthusiasm for making clear that discount usability testing is a true "crap shoot" (like playing craps at a gambling casino), I missed the boat on straight-talk.

So, now I make amends, dear reader.

I humbly pass on to you a better, clearer method of communicating our usability test results.

"Confidence Intervals" are dead.

"Margins of Error" are gone.

They are dead and gone from our client-communications, that is.

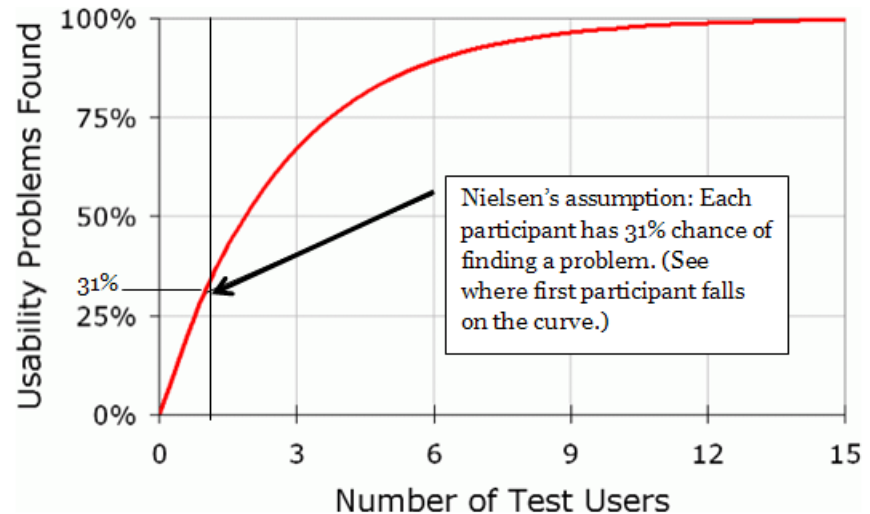
Meanwhile, keep them in mind when using the following instructions.

### Clearing the decks for straight talk

First, let's return to "first principles". I take you to that rock of usability, the picture where you may have first thought that "I can do this job".

Recall Jakob Nielsen's famous article "Why You Only Need to Test With 5 Users" published in 19 March of 2000 (as drawn from his 1993 article "A mathematical model of the finding of usability problems").

On the positive side, note that Nielson avoids talking about "pass rates". Instead, he focuses on "finding problems".



That's reasonable, since the pass rates have such a ridiculously wide margin of error (yes, there was some value in my previous article - we learned to pretty much avoid talking about "pass rates".)

So, let's talk about "finding problems".

What Nielsen showed us was that each additional testing participant gives you a chance to find more problems in your design.

Based on his 1993 data, averaged from 11 usability tests and heuristic reviews, Nielsen figured that a single participant has about a 31% chance of finding a problem. (Note that the first participant crosses the 31% "Usability Problems Found" point. I marked the spot.)

But, his chart shows you a point of diminishing returns. He suggests "5 users" finds about 85% of the problems. Well, that sounds good. But is it true today?

**Remember: the word "Ass/u/me" makes an "A\_ of U and Me"**

Here's where Jeff Sauro brings us forward a couple light years with his March 8, 2010 article "Why you only need to test with five users (explained)" First, he asks if Nielsen's 1993 findings apply to our more challenging commercial web world of 2010.

- 1) Nielsen uses an assumption which may be unrealistic for most, if not all, Web-sites. He assumes that any one test participant (or end-user) has a 31% chance of finding a problem.
- 2) Large-scale web sites (and applications) with many functions could have problems that have only a 10% chance of being found. (This is only one-third of the problem discovery potential given by Nielsen.)
- 3) If you have thousands of users, or if your application is critical (life-and-death, or high-financial impact, etc.) then you should very much be concerned with lower-frequency problems. Problems that occur even one percent of the time will bring frustration to 1,000 people if you have 100,000 users.
- 4) Last, Nielsen's approach does not apply to comparing web sites or to determining task completion rates. It does apply to discovering problems with an interface. We need to stick to that topic in our test results report.

### Talking the talk with usability test results

Here's where Jeff's discussions with his buddy James Lewis (himself a prolific writer on usability topics) pay off for us all. The following comes from Jeff's web-site article I mentioned.

**1. What Problem Occurrence Rate do you want to handle?** Early in your design work, it's OK to figure that there are problems that are easy to discover. So easy, in fact, that any one test participant has, say, a 31% chance of finding any one of them.

We'll use the 31% chance of finding a problem largely because Nielsen's chart (presented above) uses that number and you are familiar with its claims - namely that 5 participants will uncover 85% of those types of problems.

But, before we go further, let's be clear on how to say this. This is a key point of your communication with your colleagues.

We want to say "We are testing for problems that have a 31% or greater chance of being experienced by a single end-user."

Another way of saying the same thing: **"We are testing for problems that can affect 31% or more of our end-users."**

Meanwhile, remember the limitations of this statement. It means we are not finding problems that are harder to find. **We are not finding problems that affect fewer than 31% of our end-users.**

Let's be concrete about using the 31% chance of "problem occurrence".

For every 1,000 end-users, it means you are not set up to find problems that could affect less than 31% of those people. 310 people is 31% of one thousand. Less than 31% is 309 people or fewer.

Are you happy with letting those 309 people experiencing a problem on your web site or application? That is the question.

Jeff Sauro indicates that a web-site or application with thousands of users may have a rate of problem occurrence below 10%. Would you like to know how many test participants you need to find those kinds of problems?

Probably yes. So, we'll tell you the answer below. Meanwhile, we have another issue to cover first: how likely are you to actually find those problems?

**2. How serious are you about finding those problems?** Jakob Nielsen's article and chart tells us how likely you are to actually find those problems.

Check out Nielsen's statement: "After the first study with 5 users has found 85% of the usability problems, you will want to fix these problems in a redesign." The number, 85%, here means the "chance" or "likelihood" you will find those problems that 31% percent of your end-users would experience.

Are you happy with an 85% chance of finding those problems? Maybe yes, if you're in a hurry, and world peace or financial survival does not depend on your website.

In fact, Jacob Nielsen also disagrees with the 85% chance of finding those problems. Read further into his article, and you'll see that he advocates using several sets of 5 test participants (iterative testing).

He recommends that you test with 5, make improvements, and then test with another 5 participants. Note that with 10 participants, you now have a 98% chance of finding those problems. Pretty good. And with another set of 5 partici-

pants, you have raised your chances to near certainty of finding those problems (99.6% in fact).

So - realize that even Jacob Nielsen advocated more than 5 participants for testing. He recommended 15! (Do we dare tell others?)

**3. Getting serious about what problems we want to find.** I promised to return to an important topic. The question was - how happy are you with finding problems that are sufficiently popular that **at least** 31% of your end-users would find them?

The answer was probably "no", you are not happy with finding just the "popular" problems. 31% is a pretty big number. It means you are only looking for problems that have about a 1 in 3 or greater chance of popping up. Yes, we want to find these. But, what else do we want to find?

**What about problems that have only a 10% chance of being found?** They may not be real easy to find, but they remain potent in discouraging repeat visits to those 10% of your site visitors.

Jeff adds a chart to his article that gives an idea of the game-change with this new insight.

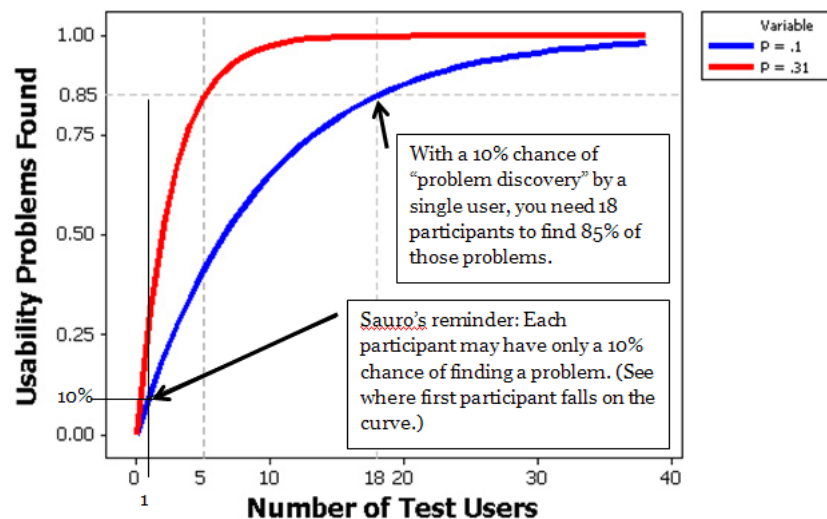


Figure: Difference in sample sizes needed to have an 85% chance of detecting a problem that affects 10% of users vs. 31% of users. You would need to plan on testing 18 users to have an 85% chance of detecting problems that affect 10% of users. (From Jeff Sauro's article.)

To help make this clear, I created a table using the formula Jeff and Nielsen used. Notice the column that shows "5" participants. You will see 84.4% in the row that shows the .31 Rate of Problem Occurrence.

This is the "85%" that Jakob Nielsen promised to find with five test participants. Notice that 10 participants lets you find 97.6% of the problems.

But, we figured that we also want to find the less popular problems - perhaps even problems with only a 10% chance of occurring.

This means that 10% of your users would experience those problems. If we want to find those problems, then 5 participants gives us only a 41% chance of finding them. To get an 85% chance of find those problems, then we need 18 participants.

Rate of Problem Occurrence	Number of participants									
	1	3	5	6	8	10	12	15	18	22
0.31	31.0%	67.1%	84.4%	89.2%	94.9%	97.6%	98.8%	99.6%	99.9%	100.0%
0.2	20.0%	48.8%	67.2%	73.8%	83.2%	89.3%	93.1%	96.5%	98.2%	99.3%
0.1	10.0%	27.1%	41.0%	46.9%	57.0%	65.1%	71.8%	79.4%	85.0%	90.2%
0.05	5.0%	14.3%	22.6%	26.5%	33.7%	40.1%	46.0%	53.7%	60.3%	67.6%
Percentage of problems found for a given Number of Participants and a given Rate of Problem Occurrence										

Question: will this make sense to your management?

If you make your pitch in the right way, it should make sense.

## Practice makes perfect

Congratulations. You made it to the end of this rather long article. I'll give you a few phrases to help you get the hang of how to talk the talk on usability testing.

**1. Avoid talking about "success rates".** We end up having to add comments about "confidence intervals" and "margins of errors" which gum up the conversation. Remember, you won't get elected president if you are too intelligent.

**2. Instead, talk about the probability of "problem occurrence".** Start with the simplest statement possible. You could even ask your management and colleagues something like "what portion of our audience can we allow to experience a problem?"

(Remember, if you are looking for problems that have a 31% chance or more of being found, then you are not looking for problems with less chance of being found. This means if 31% or more of your end-users could experience the problem, you have ignored problems that could be experienced by fewer than 31% of your end-users.)

Discuss the answers. If you have usability test data, read Jeff's article and check out his calculator that lets you see what your rate of problem occurrence has been so far.

**3. Wow. To ask these questions takes courage – not just intelligence.** It turns out you are addressing the topic of "risk mitigation". (Remember that phrase from the beginning of this article?) All business activity has its risks. Now we are asking "what is your tolerance for risk exposure?" It's that simple.

**4. Also, with your management you'll need to discuss "what kind of risk".** Discussing "kinds of risk" takes you back to your regular usability concepts of problem "severity". Categorize your problems into something like "task failure" (Severe), "task challenges" (Difficult but Doable), and "task irritations" (Glitch but Doable).

Then assign what percent of your users you are willing to let experience each type of problem. Note that some research says there is no relationship between problem severity and problem frequency. So until we learn more about this, keep each type of problem as a separate analysis.

Identify each type of problem as you analyze your testing results. For example, are you willing to let 10% of your end-users experience "task failure" (a "severe problem")? If not, pick a more stringent target (say, 5%) and select the number of subjects necessary to evaluate that alternative rate of problem occurrence.

If your results show "no severe problems" with those participants, then you know you passed your own "risk mitigation criterion".

**5. Along with with point number 2, you also have to select a target for "percent of problems found".** In our Jakob Nielsen example, Nielsen was satisfied to find 85% of the problems.

Is your team satisfied to find only 85% of the "severe" problems that cause task failure in the open-heart surgery software that you're creating? Probably not. (OK, this is an extreme case. But you get the point.)

The table I gave you above lets you pick a comfortable "percent of problems found" in relation to the number of participants and the rate of problem occurrence. Jeff's website gives details on calculating the table I showed you.

Your team may be comfortable with finding only 85% of problems during the early phases of your design work. On the other hand, prior to final release into production with 5,000 customer-service reps, you may want to show that you found 99% of the problems you set out to find.

### Recap, summary, review, shorter is better...

So you have it. Once again, with vigor...

- 1) **What rate of problem occurrence do we want to tackle?** (A 31% chance of finding a problem for a single participant also means that 31% or more of our end-users could experience the problem. Before final release, would you prefer to tackle problems that affect as few as 10% of end-users, or even 5% of end-users?)
- 2) **Consider a different rate of problem occurrence for each of the levels of severity we define,** given the cost of subjects, and penalties of failing to find the problem.
- 3) Given our commitments to #1 and #2, **what percentage of those problems do we really want to find? 85%, 95%, 99%?**
- 4) **Figure out how many participants we need.** Use my table or Jeff's calculations.
- 5) **Remember iterative design and testing.** We can distribute 15 participants across 3 test events. However, this assumes no drastic change in tasks or types of users across iterations.

### References

Sauro, Jeff, 2010. Why you only need to test with five users (explained). [www.measuringusability.com](http://www.measuringusability.com), 8 March, 2010.

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.

Simonton, D. K. (1999a). *Origins of genius: Darwinian perspectives on creativity*. Oxford: Oxford University Press.

Turner, Carl W, Lewis, James R, Nielsen, J., 2006. "Determining Usability Test Sample Size," *International Encyclopedia of Ergonomics and Human Factors*. 2nd Edition, Vol 3. (Edited by Waldemar Karwowski, CRC Press, Boca Raton, FL)

### Message from the CEO, Dr. Eric Schaffer



John, this is a great revisit to the perennial question on sample size. Clearly expectations are changing. A site which has 10% of users experiencing each of several problems would be uncompetitive by today's standards.

Also, as you indicate, with complex sites it is hard to run a single user through all the likely tasks within the normal 60 minute session (or even the 90 minute "work the participant to exhaustion" plan). So PLEASE, let's stop the nonsense with 5 person studies.

I recently had a large mobile phone company ask us to sample the Indian market with THREE participants. I asked if they wanted villagers, or Mumbaikers, or super rich. They didn't seem concerned. I declined the project.



**Human Factors**  
International

410 West Lowe, P.O. Box 2020  
Fairfield, IA 52556  
Phone: 800.242.4480  
Fax: 641.472.5412  
[hfi@humanfactors.com](mailto:hfi@humanfactors.com)  
[www.humanfactors.com](http://www.humanfactors.com)