



HFI User Experience Design Newsletter

Usability Test Reporting: “It Ain’t Over ‘Til It’s Over”

Message from the CEO, Dr. Eric Schaffer

August, 2010

[Newsletter](#)



Human Factors
International

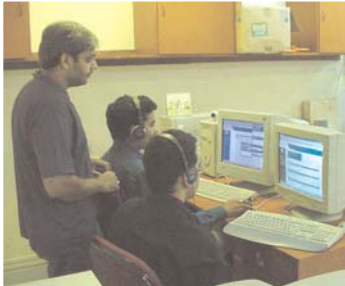
HFI User Experience Design Newsletter

August, 2010

Table of Contents

Usability Test Reporting: “It Ain’t Over ‘Til It’s Over” by John Sorflaten, PhD, CPE, CUA, HFI Writer	3
Message from the CEO, Dr. Eric Schaffer by Eric Schaffer, PhD, CPE, CUA, Founder and CEO	11

Usability Test Reporting: “It Ain’t Over ‘Til It’s Over”



Our commercial culture has amazing twists.

At Amazon.com, for a mere one cent you can purchase a book by Yogi Berra, the baseball catcher who made good with “Yogiisms” like “It ain’t over ‘til it’s over.”

This popular philosophical quote also happens to be the title of Berra’s autobiographical 1989 book covering his tectonic ups and downs in major league baseball.

Formats	Amazon Price	New from	Used from
Hardcover	—	\$7.00	\$0.01
Paperback	—	—	\$2.99
Mass Market Paperback	—	\$1.75	\$0.01

So what’s the amazing twist on a one cent book?

Well, until I wrote this article, ***I used to believe the twist was the \$3.99 shipping charge that allowed the vendor to make a few cents from even cheaper shipping.***

However, looking at the Used Book listing for Yogi’s title, I just saw that effectively, I can purchase his book for a “negative \$.96”!

In this latter case, the Amazon “Prime” price is only \$3.03—meaning the vendor gave up some of their \$3.99 shipping allowance to subsidize my purchase to the tune of 96 cents! ***Now that’s an even more amazing twist.***

So, “it ain’t over ‘til it’s over” when selling books.



What about your usability test recommendations?

So—you might think that when you write up the problems, make your design recommendations and turn them in, you’ve done your job. Right?

Wrong. You just forgot: “It ain’t over ‘til it’s over.”

The truth about truth

I’m sure you’ve had occasion to tell colleagues and managers that the source of usability problems in many cases is “ego-centric design.” HFI publicizes this truth with a famous button passed out at courses: “Know Thy Users, For They Are Not You.”

But, we might ask the same question about whether we, world-wise and experienced designers, have avoided “ego-centric re-design” when recommending solutions to usability problems.

We report the “truth” about an application’s usability faults. But how well do we frame our re-design recommendations so they are truly usable for our readers?

How truthfully can others interpret our recommendations?

This is the question asked by a trio of usability authors, each well-known in their own right: Rolf Molich, Robin Jeffries and Joe Dumas. Their 2007 study is titled “Making Usability Recommendations Useful and Usable.”

They examined how well 17 teams of usability professionals wrote up usability evaluation reports of a hotel reservation system in use by hundreds of hotels.

Each team had 1 to 5 members. Each team had an average of 1.6 persons and 5 to 40 years combined usability experience within each team. Not a bad collection of skills.

However, our trio of researchers found that 17% of the redesign recommendations from the teams **“were not useful at all.”** 19% of the redesign recommendations **“were not usable at all.”** And only 17% of the redesign recommendations **“were both useful and usable.”**

The authors report: “Quality problems include recommendations that are vague or not actionable, and ones that may not improve the overall usability of the application.”

What went wrong? How many of us thought that the hard work was the evaluation, not the write-up? Is writing usability evaluations a risky endeavor? What do you think, now?

Let’s see why “it ain’t over ‘til it’s over”, next.

Evaluating usability recommendations

We’ll jump to the crux of the study: **How did the authors evaluate the usability recommendations written by those 17 teams?** (We’ll get to other details later.)

1. The authors found 81 usability problems that were identified by at least 10 out of the 17 evaluation teams. This “consensus” among teams allowed the authors freedom from defending the definition of “usability problem.”
2. The authors developed a 5-point scale for usefulness and usability. They independently rated a “training set” of usability recommendations and then compared their ratings.

They ended their training period after reaching agreement on 89% of their trial evaluations.

3. Their evaluation scales follow. Instructions to the participants stated that recommendations should be short. I give you the essence of the authors’ comments.

The gold standard for useful and usable recommendations (give it a “5”!)

A useful recommendation—“*an effective idea for solving the usability problem.*” The description does not contain any bad suggestions.

The quality of the description is not considered as long as the idea is comprehensible “... (“quality” gets considered in the usability rating).

5—Fully useful: Meets the above criteria. “As good as it’s going to get” given a request for a short description.

A usable recommendation—“*communicates precisely and in reasonable detail what the product team should do to implement the idea behind the recommendation.*”

This follows evaluation of usefulness, but is independent of whether the idea is useful. “A recommendation that is not considered useful at all may thus still be fully usable and vice versa.”

5—Fully usable: Ditto.

Example: Issue—“**First time users find that form labels appear in the same space as the typing area for Address and Credit Card inputs. If you start to type, the labels disappear. But if you try to delete them first, nothing happens.**”

Usefulness: 5.0

Usability: 5.0

Recommendation that received these ratings: “Provide field labels next to, not within the fields... Do not present everything all on one page. Although this is a main feature of the system, it reduces the overall effectiveness by forcing too much on a limited screen space. Place the calendar and room selector on the same page, with the logic to calculate the cost (fully itemized to show taxes etc) based on selections. Use a “Proceed to booking” button to go to a second page for capturing name card info etc, and transferring the cost, room type & date information.”

Usefulness: 1.0

Usability: 5.0

Recommendation that received these ratings: “...centering the labels might allow the cursor position to be noticed more quickly”

Rater comment: “Although [this] recommendation is fully usable (there is no doubt as to what should be done), we were not convinced that centering the labels would actually improve the usability of the application.”

Would you agree on these definitions of “useful” and “usable”? It’s not unlike your Literature teacher giving you a double grade: an F for Thoughtfulness (like “usefulness”) and an A for Composition (like “usability”). This means your teacher hated the ideas, but loved your spelling, punctuation, and sentence construction!

The other ratings

4–Useful recommendation: “an effective idea [but with]... minor flaws, omissions or bad elements that may influence the usability of the resulting solution.”	4–Usable recommendation: “communicates precisely and in some detail what the product team should do. Minor details are missing; this may influence the usability of the resulting solution.”
3–Partly useful: “the recommendation also leaves significant parts of the problem unaddressed... Contains roughly equal magnitudes of good and bad ideas, when it would solve the problem only for approximately half the users, or when the idea could introduce new usability problems.”	3–Partly usable: “communicates some information about what the product team should do. The recommendation leaves some important decisions regarding the implementation... to the product team. This could introduce new usability problems in the solution.”
2–A few useful elements: “an idea that would solve only a minor part of the problem, or... only for a minor group of users.” Or, “when part of the description is so vague that the usefulness of the idea is doubtful...”	2–Mostly unusable: “most of the description is vague, unclear or difficult to understand...; Leaves many important decisions regarding the implementation of the solution to the product team.”
1–Not useful: “...might even decrease the usefulness of the product... vague or unclear...”	1–Unusable: “...totally vague, unclear, or incomprehensible for the product team.”

I think we get the idea: “3” is the middle score: half good and half, well, just plain bad. Gee. Guess we better get a better score than 3 or else get another career. Clearly, a 3 is a make or break score!

A final contrast of scores

Let's see some examples of tough love from our trio of authors...

Example:
Issue—“Credit card icons appear clickable, but they are just meant to indicate what cards are acceptable.”

Usefulness: 5.0

Usability: 4.3

Recommendation that received these ratings: “People are used to providing the card type along with the number and expiration date. It is not widely known that card type is redundant. To make this display fit the mental model of the user, it would be good if the icon reacted like buttons. These “buttons” need not work on the back side but would help the user.”

Usefulness: 5.0

Usability: 1.0

Recommendation that received these ratings: “Some users may be inclined to click on the credit card icons to specify which card they are using. Suggested solution: Change the visual presentation to discourage this unnecessary behavior”

Rater comment: “This advice is vague [and thus not usable]. Several teams were vague about how the appearance of a selected icon would be different from a non-selected one.

Usefulness: 1.3

Usability: 5.0

Recommendation that received these ratings: “The icons appear to serve no purpose. If this is the case, they should be removed so as to avoid any confusion.”

Rater comment: “...incorrect. The icons inform users of which credit cards are accepted by the hotel. On the other hand, the recommendation (“remove icons”) is very precise and actionable.”

Whew. If you can't stand the heat, get out of the kitchen, right?

Be explicit—say what you mean

The authors found that teams took seriously the challenge of offering recommendations. Teams offered suggestions 96% of the time across their 81 problems.

However, the authors found that 16% of the time teams failed to use the word “Recommendation” or the equivalent. Thus, their changes failed to capture atten-

tion. As the authors comment: “Implicit recommendations often sound like complaints or unprocessed observations of test participant difficulties...”

Recommendation: use the word “Recommendation” when making a recommendation (!) (Explicate explicitly!)

The humbling details

Our authors summarized their findings by defining “high-quality” recommendations as 4.0 and above for both usefulness and usability.

What percentage of the 81 recommendations would you guess met that modest criterion?

Well, only 17% of the 81 recommendations (14 of them) could be called both “Useful” and “Usable” or better.

The authors point out more pain among the teams. They point out that only 42% of the 81 recommendations (34) could be called both “Partly Useful” and “Partly Usable” (3.0 or better).

Can we conclude that it’s tough to write right?

Well, that’s the point of this study. Yes, it is tough to write good recommendations.

It ain’t over ‘til it’s over: how to write right

Hopefully, you’ve learned that **testing is not the hardest part. It’s the quality of your recommendations that tips the scales.**

Are you motivated, now? Here’s advice from our three authors.

1. Check your work for vague-uity.

Vague and vacuous recommendations rise when we write in a rush. Solution: have another person check for vague, unclear recommendations. Or at least read your own work *after* giving it a rest. Make sure to have clear context for your recommendations.

2. Avoid solutions that create other problems elsewhere.

Base your recommendations on data—either from usability test findings, standards that have been proven through use, or proven “best practices.” The authors advise: avoid “unsubstantiated opinions.”

3. Beware of business or technical constraints.

Show that you understand the business issues before recommending that they be rejected. For example, a logo may have a long-standing claim to a certain position on the web site. Be sure you negotiate all the pros and cons before overturning accepted practices.

4. Be sure to test sweeping changes.

You can't expect to predict all the outcomes of all your recommendations. Let your readers know ahead of time which recommendations need further testing.

5. Be specific. Be clear. This makes your writing “usable.”

Show examples to shut out vagueness. For example: writing, “change the visual presentation [of the credit card icons] to discourage unnecessary clicks,” fails to specify *what kind of change is best*. Don't expect other people to be brilliant when you can't do it yourself.

Where Yogi Berra said his thing...

The Yogiism, “It ain't over 'til it's over,” was conceived, born, and delivered on July 1973.

Yogi Berra's Mets trailed the Chicago Cubs by 9/12 games in the National League East. With Berra's inspiration, the Mets rallied to win the division title on the final day of the season.

Thank goodness for Wikipedia.

References

Molich, Rolf; Jeffries, Robin; Dumas, Joseph. 2007. Making Usability Recommendations Useful and Usable. *Journal of Usability Studies*, 2 (4), 162-179.

Yogi Berra. Downloaded 24 Aug, 2010 from Wikipedia.

Message from the CEO, Dr. Eric Schaffer



John is so right. It's NOT enough to have the correct technical understanding of a problem. We must communicate.

If we are speaking to executives we need to leave out the technical details and make the business case. If we are talking to developers we need SIMPLE explanations (NOT “Chromosteriopsis” – Say “Red text looks fuzzy”).

And, we need solid actionable design recommendations. It is pretty easy to run a test and find issues. It's a bit harder to run a systematic review and spot issues. But it's a LOT harder to do the communication and design work to make the needed improvements happen.



Human Factors
International

410 West Lowe, P.O. Box 2020
Fairfield, IA 52556
Phone: 800.242.4480
Fax: 641.472.5412
hfi@humanfactors.com
www.humanfactors.com