

Project 3: Comparison of HE and UT findings.

1. Relative higher cost-effectiveness of HE

It has been well documented that for any UTs the researchers have to spend a tremendous amount of time formulating questions and finding test subjects, setting up capturing devices and the like. This was no exception for our UT as well. It involved more than 20 hours of formulating both pre-test questionnaires, tasks for the UT, post test interview, conducting the tests and analyzing the test results. By comparison, the HE took less than half the time.

2. Convergence of results

As stated above there was a degree of convergence between the UT and HE that predominantly centered on issues of feedback efficiency and in turn how this feedback could be counter-productive or just contrary to real world standards. Also there is another issue at play here, and that is the interpretation of the test data during UT could be flawed based on subjective parameters disguised as tools for the test.

3. Accuracy and objectivity of UT results and misidentification of problems in HE

UT and HE are needed in different moments during the design process and we believe that within this project they should have been done in reverse. Specifically, the UT that we conducted would have been far more effective if we had HE data from the initial design process. In this regard our UTs would have been far more focused if that had been made available.

Our findings that the UT results are richer and more complete than our HE results might have been influenced by our lack of expertise in the field of HE. The nature of Usability Testing is that we follow a set of proscribed rules and methodologies. They are laid out in a manner which leaves less margin for error as compared to HE. The field of interpretation of the 10 rules of HE (3) is very wide. (See insight paper for discussion regarding this matter).

4. Linking intrinsic feature to payoff performance

We agree with the finding that HE identifies intrinsic features without attempting to link the feature to payoff or cost of performance. HE as a UEM is not indicated in predicting payoffs. Rather as stated below in item 8, HE is intended to provide overall perceptions of the given tool.

5. Pool of evaluators vs. population of testers

Law & Hvannberg found that it was less than valid to have the same researcher conducting the UT and the HE. This was despite the fact that our order of tests was reversed – we conducted the UT first before the HE. We found that it was difficult to separate ourselves from the previous UT findings and view the application through the lens of Heuristic Evaluation.

We also agree with L&H that the ability to find a large and non-self-selecting pool of volunteers increases the validity of UT results. Our results from the UT were a reflective of a broad user base. We are not experienced HE evaluators and we do not have domain knowledge. Nielsen (1) found that double experts, evaluators with both domain knowledge and evaluative skills do better than ‘single’ experts. As graduate students in usability we probably enter below the ‘single’ expert level of expertise. Thus, our identification of HE problems is most likely less than complete.

6. Positive findings

In our UT project we found that the Address Book is easy to use if the user is searching for someone known to him. This was a positive finding that was not captured anywhere in our HE findings.

So we agree with Law & Hvannberg that HE as a usability evaluation method does not evaluate the positive aspects of a given application. It seems that the entire conception of HE is designed to frame the evaluator to capture errors. While capturing errors is certainly a key task of usability evaluation, ignoring positive findings would be a mistake. When recommending changes, a usability evaluation expert cannot afford to forget the strengths of an existing application. Doing so might result in recommendations for changes that might cause existing strengths to be diluted or worse, discarded.

7. Predictive power of UEMs

The Law & Hvannberg paper considers this question in light of how predictive HE is of the problems uncovered in UT. In our exercise, we reversed the order of events and conducted the UT before the HE. We found, as mentioned in item 5 above, that it was difficult to separate our conclusions in the UT exercise from our HE findings. We found substantial overlap in our findings between each UEM.

A related issue is that the L&H paper recommends the conducting of HE and the implementation of those changes before UT is even attempted. In the light of this recommendation, the predictive power of HE is rarely tested in the standard progressions of an evaluation tests.

8. Accumulative insights into problems

Our HE findings also seemed to identify the problems in more general terms. Attempts at identifying specifics of the issue led to problems identifying if the specific issue belonged to one heuristic or to another. For example, one evaluator stated that “the reply after a search error is vague and cryptic, no information given on the type of error” belongs under the category of “**Help users recognize, diagnose, and recover from errors**”. Refining the aspects of the problem, for example, ‘error message does not repeat search terms’ causes the problem to fall under another heuristic as well; “**Recognition rather than recall.**”

HE as a UEM seems to provide the leeway for researchers to step back and assess a given usability problem against an independently (arguably rigorous) template. Each heuristic is based on the multiplication of several independent problem areas. It might be described as providing an insight into the *experience* of using an application as opposed to the *problems* of using an application.

Reflection on UT and HE

In reflection, we feel that the usability testing and heuristic evaluation were performed in the incorrect order. It might have been easier to develop better testing scenarios and procedures had we done our heuristic evaluations prior to writing the format of the usability tests. Also, it would have been interesting to see how our own heuristic evaluations might have differed had we not already administered a number of usability tests on the interface. One can attempt to disregard their prior knowledge of a system’s errors when dealing with such a situation; however, the usability tests provided us with a number of usability problems that seemed inevitable in our heuristic evaluation.

We also failed to perform pilot tests on our initial usability tests which would have provided us with better testing procedures and results. The pilot tests would have also provided us with invaluable information concerning the hardware testing we should have performed prior to the actual usability tests. We ran into a few difficulties throughout our usability tests with the dual software packages overloading the computer we were working on.

Overall, however, we feel that the experience we received from performing both the usability tests and the heuristic evaluations were invaluable to the advancement of our knowledge of experience testing. Being provided with a real-world situation such as the testing of an Indiana University system is an opportunity that we are grateful to have had.

References:

1. Nielsen, Jakob. Finding usability problems through heuristic evaluation. Proc. CHI1992, ACM Press (1992),373-380.
2. Law, Lai-Chong & Hvannberg, Ebba Thora . Complementarity and Convergence of Heuristic Evaluation and Usability Test: A Case Study of UNIVERSAL Brokerage Platform. NordiCHI (2002) 71-80
3. Ten Usability Heuristics. Source:
http://www.useit.com/papers/heuristic/heuristic_list.html