The Metaphors-of-Human-Thinking Technique for Usability Evaluation Compared to Heuristic Evaluation and Cognitive Walkthrough

Erik Frøkjær & Kasper Hornbæk

Datalogisk Institut & Naturvidenskabeligt ICT Kompetencecenter, Københavns Universitet, erikf@diku.dk & khornbaek@nik.ku.dk

Introduction

This paper reports results from two comparative studies where the metaphors-of-human-thinking usability evaluation technique, MOT (Hornbæk & Frøkjær 2002; Frøkjær & Hornbæk 2002a) are compared to heuristic evaluation, HE (Molich & Nielsen 1990) and cognitive walkthrough, CW (Wharton et al. 1994), and thereby follow up on our paper on the utility of Naur/James inspired psychology in HCI (Frøkjær & Hornbæk 2002b) presented at last years symposium.

HE Compared to MOT

To understand the effectiveness of metaphors of human thinking as a usability inspection technique, we conducted an experiment comparing MOT to heuristic evaluation (Hornbæk & Frøkjær 2003). Eighty-seven computer science students used either HE or MOT to evaluate a web application (http://punkt.ku.dk). Each subject individually performed the evaluation supported by scenarios made available by the developers of the web application. Forty-four subjects received as description of MOT a pseudonymized version of (Hornbæk & Frøkjær 2002); 43 subjects received a description of HE from (Nielsen 1993, 19-20 and 115-163). In all, subjects identified 911 problems.

In order to find problems that are similar to each other, we undertook a consolidation of the problems. In this consolidation, the two authors grouped together problems perceived alike. This resulted in a list of 341 consolidated problems. Next the client (i.e. the person who manages the development of the web application and is responsible for developing the design) assessed each consolidated problem. We asked the client to assess for each consolidated problem: severity (on a scale from 1 to 3), if design ideas were gotten from the problems (yes or no), if the problem was novel

1

	HE (N=43)	MOT (44)
Number of problems	11.3 (6.2)	9.6 (5.7)
Severity ***	2.4 (0.9)	2.2 (0.5)
Complexity ***	3.2 (1.0)	3.00 (0.8)
Novel problems***	3.8 (2.8)	2.0 (1.5)
Design ideas	2.5 (1.9)	2.2 (2.2)

Table 1: Usability problems identified with heuristic evaluation (HE) metaphors of human thinking (MOT). Severity was graded 1,2 or 3, where 1 was given to a very critical problem and 3 was given to cosmetic problem. Complexity was graded from 1 to 4, where 1 was given to a very complex problem and 4 to a simple problem. All other rows refer to the average number of problems found by a subject. ***=significant difference between techniques.

(yes or no), and the perceived complexity of solving the problem (on a scale from 1 to 4).

Table 1 shows some results from the experiment. By analysis of variance, we find no difference between the number of problems subjects identified with the two techniques, F(1,85)=1.76 p>.1.

Analyzing the client's assessment of the severity of problems, we find a significant difference between techniques, F(1,85)=15.51, p<.001. The client assesses problems identified with MOT as more severe (M=2.21; <u>SD</u>=0.73) than problems found by HE (M=2.42; <u>SD</u>=0.87).

The complexity of the problems identified is significantly different between techniques, F(1,85)=12.94, p<.001. The client assesses problems found with MOT as more complex to solve (<u>M</u>=3.00, <u>SD</u>=0.80) compared to those found by HE (<u>M</u>=3.21, <u>SD</u>=0.96).

Concerning the number of novel problems, HE identifies significantly more than MOT does, F(1,85)=14.59, p<.001. For both techniques, novel problems on the average are less severe ($\underline{M} = 2.31$; $\underline{SD} = 0.75$), are less complex ($\underline{M} = 3.48$; $\underline{SD} = 0.71$), and 41% are only found by one subject, suggesting that novel problems are mostly cosmetic and somewhat esoteric problems.

For reading and performing the inspections, the subjects reported spending for MOT on average 4.0 hours ($\underline{SD}=2.3$) and for HE 5.8 hours ($\underline{SD}=3.8$).

2

This difference is significant and large (Mann-Whitney U=546.5, z=-2.88, p<.01).

CW Compared to MOT

The aim of this second study is to compare the effectiveness of inspection by metaphors of human thinking with cognitive walkthrough. As a supplement to quantitative data from the evaluations, participants are required to keep a diary during the evaluation to shed light on problems and insights experienced when using the techniques. Data from the experiment will help improve MOT and CW, and identify strengths and weaknesses of the techniques. 20 participants evaluate and redesign web sites using MOT and CW. Each of the techniques was used to evaluate and redesign an ecommerce web site. The site evaluated in the first week was http://www.gevalia.com; in the second week http://www.jcrew.com. Both sites are included in a large professional study of e-commerce sites (Nielsen et al. 2001), which offers insights into usability problems of e-commerce sites. Below, we give initial quantitative results of the study; the qualitative results will be reported later.

Analysis of variance show that participants identify significantly more problems using MOT compared to CW, F(1,19)=8.68, p<0.001. On average, participants identify 11.8 (SD=7.52) with MOT and 9.0 (SD=8.18) problems with CW, that is 31% more. In raw numbers, 13 participants find more problems with MOT, 3 identify the same number of problems, and 4 identify more with CW.

We find no difference in the severity ratings assigned by participants to the usability problems, F(1, 19)=3.35, p>.05. On the average participants using MOT assess the severity of the problems as 2.31 (<u>SD</u>=.72); using CW average severity is 2.25 (<u>SD</u>=.69).

We compared the usability problems found by participants to a reference collection of usability problems of particular relevance to e-commerce web sites (Nielsen et al. 2001). Both techniques succeed in finding problems that hit the reference collection. In combination the two techniques achieve 51% coverage of the collection (note that only two of the web sites studied in Nielsen et al. 2001 were used here).

Using MOT, participants identify usability problems covering a broader group of problems in the reference collection, F(1,19)=4.48, p<.05. Among all evaluators, MOT identifies 36 problems (17%) in the reference collection that CW did not find; CW finds only 21 problems (10%) in the reference collection that MOT did not find.

3

These two experiments show that the metaphor-of-human-thinking technique can be an effective and convenient alternative or supplement to the two well-known usability inspection techniques, heuristic evaluation and cognitive walkthrough.

References

Frøkjær, E. & Hornbæk, K. (2002a), "Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing", *In Proceedings of HF2002/OzCHI 2002*.

Frøkjær, E. & Hornbæk, K. (2002b), "Studying the Utility of Metaphors of Human Thinking in HCI", *Proceedings of the Second Danish Human-Computer Interaction Research Symposium. DIKU Technical Report no. 2002/19*, 19-21.

Hornbæk, K. & Frøkjær, E. (2002), "Evaluating User Interfaces with Metaphors of Human Thinking", *In Proceedings of ERCIM Workshop on User Interfaces for All Proceedings of 7th ERCIM Workshop "User Interfaces for All"*. Also in N. Carbonell & C. Stephanidis (2003) "Universal Access", *Lecture Notes in Computer Science 2615*, 486-507, Berlin: Springer-Verlag.

Hornbæk, K. & Frøkjær, E. (2003), "Metaphors of Human Thinking: A new Tool in User Interface Design and Evaluation", *In Proceedings of Interact 2003 IFIP TC13 International Conference on Human-Computer Interaction*, 781-784.

Molich, R. & Nielsen, J. (1990), "Improving a Human-Computer Dialogue", *Communications of the ACM*, 33, 3, 338-348.

Nielsen, J. (1993), Usability Engineering, San Diego CA: Academic Press.

Nielsen, J., Molich, R., Snyder, C. & Farrell, S. (2001), *E-commerce User Experience*, Fremont CA: Nielsen Norman Group.

Wharton, C., Rieman, J., Lewis, C., & Polson, P., (1994), "The cognitive walkthrough method: a practitioner's guide", 105-140, *in* Nielsen, Jakob and Mack, R. L., *Usability Inspection Methods,* John Wiley & Sons.