

Individual and pair performance in usability tests

Eija Suikola

Teamware Group

P.O. Box 135

FIN-00381 Helsinki

Finland

+358 9 5128 2660

eija.suikola@teamware.com

1. INTRODUCTION

Usability tests with a pair of participants have often been recommended to alleviate problems in thinking aloud in single-participant tests. HCI community has been relatively immune to results in distributed cognition indicating that in some situations “two heads are better than one.” This would mean that results in usability tests with a pair do not give comparable results to results by individual users.

Experience in usability tests has indicated that thinking aloud method (Ericsson & Simon, 1984) has limitations. Many users find it difficult to think aloud. Some researchers and practitioners in HCI (e.g. Kennedy, 1989) have recommended usability tests with pairs to remove the problems in producing verbal accounts. However, HCI people have not questioned the validity of this recommendation with the exception of (Hackman and Biers, 1992). It is not self-evident that the results of single-user testing and pair testing should be equal.

In distributed cognition, researchers have claimed that the problem solving of individual participants and pairs is different measured by performance and outcome of problem solving (Flor & Hutchins, 1993; Miyake, 1986). Use of software is cognitively speaking problem solving. Based on these results we can raise the question whether results of pair tests actually give valid results about usability of software that is intended to be used by individual users working alone.

Both the above mentioned cases studied only pairs and compared the results to previous studies or theories about individual problem solving. The benefits of pairs have been questioned by other researchers in distributed cognition and contradicting results have been found (e.g. Andersson & Rönnerberg, 1995). According to them two heads are not always better than one.

2. THE STUDY

All material on each page should fit within a The reported research tested the comparability of individual-user and pair usability testing by comparing the problem solving of users of a software program. The study was done from the viewpoint of distributed cognition. The research questions focused on two areas. First, are there differences in the process and results of problem-solving performance of individual users and pairs performing the same tasks? And second, are there differences in the externalisation of cognitive processes?

2.1 Methods

Six individuals and six pairs performed the same two tasks in similar video taped conditions in a usability lab. Participants used a groupware calendar system, TeamWARE Calendar™. The tasks required users to find free time (task A) and specify criteria for a repeating appointment (task B). After a rehearsal task participants performed the test tasks. The order of the tasks was varied across test sessions.

Behavioural and verbalisation data were analysed using several different variables and analyses methods. In this paper, only a part of main results are discussed. The behavioural data were analysed using performance time, number and type of errors, error management strategy.

A content analysis was done on the verbal protocols. The protocols were divided into propositions. Propositions were classified into three categories: *reflection* (propositions externalising participants' problem solving), *commenting* (propositions restating the observed behaviour), and *other* (propositions that did not fit either of the above classes, left out of analysis).

A statement like “do I have to do it manually / or could there be something that helps to do it ... like automatically” was categorised as a reflection. It illustrates the participant's problem solving when

planning how to define recurring dates. A statements like “and then action for days off / ok, like this” that shadows the user’s actions while viewing a dialog and selecting options was categorised as commenting.

2.2 Results

The results did not indicate statistically significant differences between individuals and pairs in performance time analysed by task type and the order of the tasks, frequency or type of errors, total amount of propositions analysed by task type and the order of the tasks.

However, some noteworthy observations were made showing possible behavioural tendencies. The performance time of pairs decreased in the second task more than individuals’ performance time. Both individual users and pairs verbalised less in the second task. The decrease was bigger in pairs.

A closer analysis of the participants’ behaviour indicated that on the whole the performance differences were better explained by the widely varying effectiveness of error management strategies. Some individuals outperformed some pairs and therefore their total performance was better.

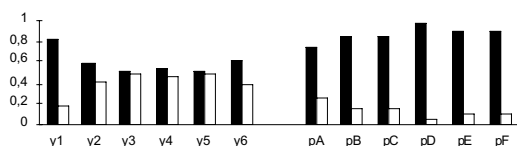


Figure 1. Proportion of reflection (black) and commenting (white) propositions in task A. Individuals = y1-y6 and pairs = pA-pF.

The content of verbalisation indicated differences between individual users and pairs (Figure 1). Pairs verbalised statistically significantly more reflection and less commenting than individuals did. Working in pairs encourages externalisation of one’s thoughts to the partner (reflection). However, users working alone simply describe their behaviour (commenting) instead of externalising their problem solving.

3. DISCUSSION

Results did not indicate group-level differences in performance between individuals and pairs, though results indicated that pairs’ performance and verbalisation

changed more in the second task than individuals’ performance and verbalisation. These changes may indicate a stronger learning effect on pairs’ behaviour. Another possible explanation is that during the first task pairs’ performance suffered from co-ordination difficulties.

Pairs’ verbal protocols were richer in content and more informative of the cognitive processes related to problem solving during task performance. However, the problem solving process itself and its results were not improved by having a pair to solve a problem. These results support Hackman & Biers (1992) and results in distributed cognition that claim that pairs are not always better than individuals.

Total performance was better explained by idiosyncratic differences in error management than the mode of working (alone or in pairs). Pairs may, in fact, be used in usability testing without the risk of getting over-positive performance results.

4. REFERENCES

- Andersson, J. & Rönnerberg, J. Recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology*, 9 (1995), 199-211.
- Ericsson, K.A. & Simon, H.A. *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge, MA. 1984.
- Flor, N.V. & Hutchins, E.L. Analyzing distributed cognition in software teams: A case study of team programming during perfective software maintenance. In R.M. Baecker (Ed). *Groupware and computer-supported cooperative work. Assisting human-human collaboration*. 1993. Morgan Kaufman, 272-286.
- Hackman, G.S. & Biers, D.W. Team usability testing: Are two heads better than one? *Proceedings of the Human Factors Society 36th Annual Meeting*, October 12-16, 1992. Atlanta, Georgia/Innovations for Interactions. Human Factors & Ergonomics Society, 1205-1209.
- Kennedy, S. Using video in the BNR usability laboratory. *SIGCHI Bulletin*, 21,2 (1989), 92-95.
- Miyake, N. Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10 (1986), 151-177.