Effects of Social Facilitation & Electronic Monitoring on Usability Testing

Erin J. Harris
Southern Illinois University Edwardsville

Jerry B. Weinberg
Southern Illinois University Edwardsville

Susan L. Thomas
Southern Illinois University Edwardsville

Danica Geaslin
Southern Illinois University Edwardsville

PAPER SUMMARY

Topic Category:	Usability method implementation or adaptation Issues and strategies for experienced usability professionals
Presentation Type:	Overview of a concept, philosophy or methodology
Intended Audience:	Basics for People Who Are New to Usability Topics for People Who Are Experienced in Usability but New to the Topic

YOUR BACKGROUND IN THIS MATERIAL

This study was conducted as a funded research project in the Undergraduate Research Academy of the university. The project was performed under the guidance of professors in the Departments of Computer Science and Psychology. I hold an honors degree with a major in computer science and a minor in psychology. As an undergraduate honors student I have been involved in studies of usability testing and the development of the Computer Science Department HCI curriculum. I am currently conducting graduate level research work in HCI. The early findings of this study were presented to usability professionals in the St. Louis area at an ACM SIGCHI meeting.

ABSTRACT

Social facilitation theory states that a person's performance is affected when being observed. Aspects of usability testing were studied to determine the impact of social facilitation. Subjects were run through a usability test varying the amount of observation. Statistically significant effects were found in the number of errors made.

INTRODUCTION

Usability Testing has the components of two social-psychological phenomena that may impact the results: (1) social facilitation through direct observation and (2) the effect of electronic performance monitoring (EPM).

EPM utilizes a wide range of technological practices often used by employers to monitor their employees, such as telephone and computer monitoring and videotaping. There appears to be a link between EPM and both increased levels of stress and decreased levels of productivity and work quality [1].

A possible explanation for the effects of EPM is social facilitation. Social facilitation theory states that the mere presence of others will affect a person's performance [5]. As originally proposed, social facilitation was defined as improvements in performance produced by the mere presence of others [3, 17]. However, some research has shown a decrease in performance for tasks that are unfamiliar [14]. Drive Theory of Social Facilitation offers an explanation for these inconsistent findings [20]. According to the theory, the dominant response for the situation is intensified when others are present due to arousal; thus, performance can be enhanced or impaired.

Subsequent theories attempt to explain the different impact arousal can have when one is in the presence of others: (1) alert and vigilant, (2) apprehensive about evaluation, and (3)

distracted. While all three theories may explain the negative effects of EPM, apprehension about evaluation is the most relevant. It is important to note that evaluation apprehension can occur even when the purpose of monitoring is not for the evaluation of the person's performance. Being monitored for any reason, such as usability testing, can make a person apprehensive.

The link between EPM and increased levels of stress has been demonstrated in numerous studies. For example, in a study involving 762 telecommunications workers, those that were monitored reported "higher levels of tension, anxiety, depression, anger, and fatigue" [17, p. 23]. Other studies have found that participants working on tasks requiring much thought and skill performed significantly worse when under EPM [2]. However, the performance of participants working on simple repetitive tasks was increased when electronically monitored [1].

Support for the enhancement of dominant responses in the presence of others has been found in such diverse areas as cockroaches running a maze and pool players' shot-making ability. Cockroaches were able to complete a simple maze much faster when in the presence of other cockroaches. Conversely, they had more difficulty completing the complex maze when in the presence of other cockroaches [21]. In the same way, good pool players increased their shot-making ability when observed while less talented players dropped their percentage of made shots [12].

The enhancement of the dominant response is intensified when people believe that the observer is assessing their performance. For example, while blindfolded observers did not affect autoworker's job performances, observers who were in a position to see the performance elicited a dominant response – good workers performed better and poor workers performed worse [6]. This effect can be even more pronounced if the observer is seen as being an expert [10].

METHODS

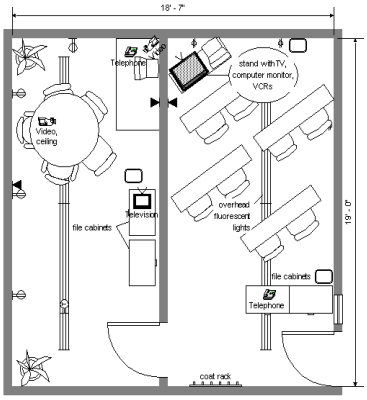
Hypothesis

Given that the basic set-up of usability testing closely resembles EPM, this study was designed to determine whether the known effects of EPM and social facilitation carry over to usability testing in the software design process. It can be hypothesized that individuals performing usability tasks with the presence of a facilitator will exhibit enhanced performance when completing a simple task and impaired performance when completing a complex task. These effects should be magnified when the evaluation component is made more salient through a constant reminder of the electronic monitoring or through meeting the software design team.

Procedure

The study examined three aspects of Usability Testing: (1) persistent knowledge of being observed, (2) presence of a facilitator, and (3) persistent knowledge of being electronically monitored. Four groups, each consisting of 25 randomly assigned participants, were asked to perform a usability test on Microsoft Word. Each group differed in the way the usability test was conducted in order to control each of the aspects of the study. The subjects were recruited from the introductory computer science programming class. This is a class for both computer science majors and non-majors. Further, Microsoft Word is the primary word processing application available to students across the university. Therefore, all subjects were computer literate and somewhat familiar with Microsoft Word.

User Interaction Room



Observation Room

Figure 1. User Interaction and Observation Rooms

The Usability Lab where the research was conducted is a split room set-up to support Usability Testing (See Figure 1). One side of the lab provides an area for the user and facilitator to perform computer tasks. The other side provides a monitor and video tape recorder for observation.

All subjects were instructed that they were participating in a software design test to determine the usability of a particular software application. All subjects were also informed that they were being videotaped and that the videos were to be viewed only for the purpose of collecting information about the ease of use of the software, and would be erased immediately after the study. All subjects were asked to complete the same 8 simple and 8 complex tasks using Microsoft Word (See Figure 2). Similarly, all subjects were instructed to talk aloud.

Simple Tasks

- Center and underline the title of the document.
- Insert a page break at the end of the document

Complex Tasks

- Add a center tab stop at the 3-inch mark from the left side of the screen.
- When you add a header to Microsoft Word, the header is added to all pages of the document. Setup Word so that the header of the first page of the document is different from the rest of the pages.

Figure 2. Examples of Simple and Complex Tasks

The subjects were instructed to continue working on the task until the task was completed or a time limit expired. The completeness and number of errors for each task were recorded, where an error is described as any divergence from the expected behavior [16]. For example, choosing menu items that are not part of the task's solution would constitute an error. Each subject's run was captured using screen recording software for the purposes of counting errors. The recording software can run in a background mode, so as not to remind the user of the recording.

The first of the four subject groups, termed "Facilitator Group", served as the control group for the study. Subjects in this group performed the usability test in the presence of a facilitator who was seated behind and just off to the side of the subject so as to have a view of the screen. The facilitator occasionally reminded the subjects to talk out loud as they worked and indicated when the subjects could go on to the next task (either when a task was completed or when the time for the task elapsed).

The second group, termed "No Facilitator Group", was run without a facilitator present in the room with the user. As with all subjects, they were informed that videotaping would be done and that it is important that they talk out loud. The only communication with the subject was a beep through a two-way radio to indicate the task was completed or the time on a task had expired and to go on to the next task.

The third group, termed "Facilitator + Reminding Group", included the presence of a facilitator as in the Facilitator Group, but added a reminder of the electronic monitoring through frequent communication with the software design team who was watching via closed circuit television. To simulate the presence of a design team, an experimental cohort was stationed in the Observation Room (see Figure 1). The communication was done by means of two-way radios. The nature of the communication was to simulate gathering information about how a task was performed. The questions were phrased similarly, for example "Could you tell us why you did it this way" or "Did you know you could do it using this other menu" or "Why do you think you had difficulty completing this task?". In addition, the recording software was set to flash a bounding box around the area being recorded, in this case entire desktop.

The fourth group, termed "Facilitator + Design Team", included the presence of a facilitator as in the Facilitator Group, but subjects were also introduced to the "software design team" prior to the usability test. Subjects were first led into the Observation Room where they met three to four members of the design team. In addition, they were shown the closed circuit television displaying the view the team would be observing. The television shows a split screen picture, one camera focused on the subject and the second focused on the computer screen that the subject is using. Various user models were on the white board walls of the Observation Room. A group of confederates was used to simulate the software design team. Members greeted the subject and introduced themselves by name.

Group	Treatment
Facilitator (Control Group)	With presence of a facilitator
No Facilitator	Without presence of a facilitator
Facilitator + Reminding	Frequent communication with the software design team
Facilitator + Design Team	Introduced to a group of confederates and shown the monitoring room prior to the Usability Test

Table 1. Experimental Groups

RESULTS

In order to examine the effects of social facilitation and electronic monitoring, two 4 × 2 mixed Analysis of Variances (ANOVAs) were conducted with group (No Facilitator vs. Facilitator vs. Facilitator + Reminding vs. Facilitator + Design Team) as the between subject variable and type of task (simple vs. complex) as the within subject variable. Errors and task completion were the dependent measures. The mean number of errors committed by each subject group is shown in Figure 3.

The ANOVA for errors revealed a main effect for the type of task. Fewer errors were made on simple tasks (M = 9.74) than on the complex tasks (M = 53.2), $F(1, 96) = 367.34$, $p = .0001$.

A main effect was also found for group. The No Facilitator group (M = 27.98) and the Facilitator + Constant Reminding group (M = 26.10) performed similarly and had

significantly fewer errors than the Facilitator group (M = 35.82) and the Facilitator + Design Team group (M = 35.98), which also performed similarly, F(3, 96) = 3.50, p = .018.

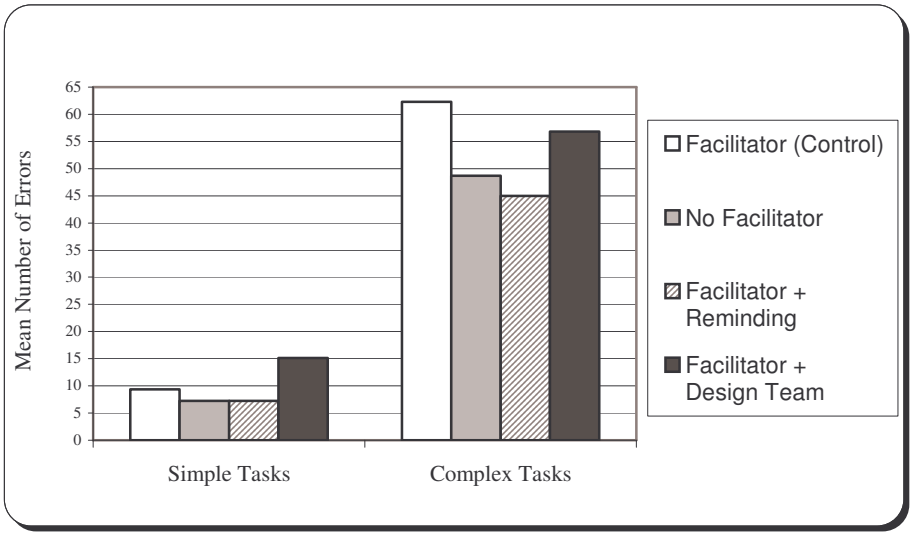


Figure 3. Task Errors

The interaction between group and the type of task did not reach statistical significance, F(3, 96) = 2.10, p = .106.

The ANOVA for task completion revealed a main effect for the type of task. A larger number of simple tasks were completed (M = 7.6) than complex tasks (M = 5.92), F(1, 96) = 162.04, p = .0001. There was neither a significant main effect for group nor a significant interaction between group and the type of task (p's > .4).

Discussion

Social facilitation theory predicts that a person's dominant response to performing a task is enhanced: performance on simple tasks improves and performance on complex tasks degrades. Our experimental results show that the effects of social facilitation are present when conducting a usability test. Specifically, in the presence of a facilitator, users performed significantly worse, or made a greater number of errors, on complex tasks than when no facilitator was present (No Facilitator group vs. Facilitator and Facilitator + Design Team groups).

Contrary to our hypothesis, the Facilitator and Facilitator + Design Team groups also experienced impaired performance by making a greater number of errors on simple tasks when compared with the No Facilitator group. The explanation for this can be found by more closely examining the simple tasks used in this experiment. In order for performance to be enhanced due to social facilitation, the correct response to the task being completed must be well learned, repetitive, or instinctive. In this study, the simple tasks, while easier to complete than the complex tasks, were not so well learned that performance enhancement could be achieved in the presence of a facilitator and design team.

The Facilitator + Reminding group performed very similarly to the No Facilitator group in the number of errors made for both complex and simple tasks. One explanation of this is the nature of the constant reminding. For each subject, a confederate on a two-way radio would occasionally ask the facilitator to remind the user to talk out loud, ask the user why they had difficulty with a task, and occasionally ask them why they had not considered a particular menu or dialog box that would have led to a solution. A frequent response to this last interaction would be "I wasn't paying attention" or "I just wasn't thinking". This type of

constant reminding appeared to result in the user paying closer attention to the menus and complex dialog boxes he or she encountered, which inhibited the social facilitation response and enhanced performance. Wright & Converse (1992) had similar findings when comparing the error frequency of users performing usability tests silently versus users instructed to talk aloud and provide explanations for their actions. Those users instructed to provide an explanation made significantly fewer errors than users in the silent condition. While these subjects committed fewer errors, Wright & Converse note that this method for usability testing could cause a bias towards hiding software difficulties [19].

Similar to the results for errors, the type of task greatly impacted the number of tasks completed. Specifically, a significantly greater number of simple tasks were completed than complex tasks. However, task completion did not follow the same trend as errors when looking at groups. All groups completed nearly the same number of both simple and complex tasks. Since a time limit was used to move users on to the next task, a ceiling effect was experienced for task completion, resulting in very little difference between the groups.

Conclusion

The practice of usability testing hinges on users making mistakes in the presence of a software designer or design team. The design team uses the results of the test to suggest changes or refinements to the design of an application or suggest design ideas for new applications. It is important to recognize that usability testing takes place within a social-psychological context, and this context can have an impact on the results of the test. Medical physicians have recognized a similar problem when taking blood pressure measurements to diagnose hypertension. Specifically, a patient's blood pressure can rise just from the psychological effects of the office visit and having their blood pressure tested [11]. This is termed the "white coat response". Our results clearly indicate that the context of the testing environment can have a significant impact on the user's performance. This in turn can impact the design team's conclusions about the interaction design of their applications. This could be particularly detrimental if only a few users are available to do testing.

Our results show that the particular procedures for conducting usability testing can have a significant impact on the outcome of the test. Specifically, the results appear to uphold the idea of minimizing the intrusion of the testing procedure into the user's awareness. While our experimental set-up was designed to re-create conditions that arise in Usability Labs, work reported by Robinson-Staveley & Cooper [15] indicates that even the mere presence of another person can affect performance on a computer task. Additionally, the affect of social context can be further intensified by the gender mix of user and facilitator.

Figure 4. Camera placement in the Usability Testing Lab

The results suggest that at least some of the usability testing situations should be conducted with minimal interaction as possible. Further, when setting up a Usability Lab, one should try to be sure that all such observation and monitoring equipment is placed in a way that is not constantly reminding the user of the observation. The cameras and microphones in our usability testing room are mounted in the ceiling to be unobtrusive (See Figure 4). Though we did not have a two-way mirror in our lab to test with, our results

imply that labs with large two-way mirrors could have an impact on the user's awareness of being observed. Further study is needed to clarify this point.

Our results clearly indicate that the social-psychological context of the usability test can have a significant impact on a user's performance. While we only measured performance, conducting experiments with different lab set-ups, gender differences between the facilitator and users, and different introductions into the process while measuring a person's state anxiety could provide a clearer picture as to the amount of anxiety each of these factors contribute. Alternative methods of conducting usability tests should be explored too, such as a think after method where the user conducts a usability test in private while running a screen recording application in the background, making the user completely unaware of the observation. A post interview could then be conducted with the user while reviewing the recording to determine what the user may have been thinking at the time [13].

ACKNOWLEDGEMENTS

We would like to thank the Undergraduate Research Academy of Southern Illinois University Edwardsville for funding this research.

REFERENCES

1. Aiello, J.R. & Kolb, K.J. (1995). *Electronic Performance Monitoring: A Risk Factor for Workplace Stress*, pages 163-179. In S. Sauter & L. Murphy (Eds.), Organizational Risk Factors for Job Stress. Washington, DC: American Psychological Association.
2. Aiello, J. R., & Shao, Y. (1993). *Electronic Performance Monitoring and Stress: The role of feedback and goal setting*. In M. J. Smith & G. Salvendy, Eds., Human-Computer Interaction: Applications and Case Studies, Elsevier-Science, 1011 - 1016.
3. Allport, F. H. (1920). *The Influence of the Group Upon Association and Thought*. Experimental Psychology, 3, 159-182.
4. Beyer, H., and Holtzblatt, K. (1997). Contextual Design: A customer-centered approach to systems designs. San Francisco, CA: Morgan Kaufman Publishers.
5. Chomiak, A.A., Kolb, K.J., & Aiello, J.R. (April 1993). *Effect of Computer Monitoring and Distraction on Task Performance*. Paper presented at the meeting of the Eastern Psychological Association, Washington, D.C.
6. Cottrell, N.B., Wack, D.L., Sekerak, G.J., & Rittle, R. H. (1968). *Social Facilitation of Dominant Responses by the Presence of Audience and the Mere Presence of Others*. Journal of Personality and Social Psychology, 9, 245-250.
7. Geen, R. G. (1989). *Alternative Conceptions of Social Facilitation*. In P. B. Paulus (Ed.), Psychology of Group Influence (2nd ed. pp. 15-51). Erlbaum, Hillsdale, NJ.
8. Gentner, D. & Grudin, J. (1996). Design *Models for Computer-Human Interfaces*. IEEE Computer, 29, 6, 28-35.
9. Guerin, Bernard. (1993). Social Facilitation. Cambridge University Press, Great Britain.
10. Henchy, T. & Glass, D.C. (1968). *Evaluation Apprehension and the Social Facilitation of Dominant and Subordinate Responses*. Journal of Personality and Social Psychology, 10, 446-454.
11. MacDonald, M.B., Laing, G.P., Wilson, M.P., & Wilson, T.W. (1999). *Prevalence and Predictors of White-Coat Response in Patients with Treated Hypertension*. Canadian Medical Association Journal, 161, 265-269.
12. Michaels, J.W., Blommel, J.M., Brocato, R.M., Linkous, R.A., & Rowe, J.S. (1982). *Social Facilitation and Inhibition in a Natural Setting*. Replications in Social Psychology, 2, 21-24.

13. Ohnemus, K. R. & Biers, D. W. (1993). *Retrospective Versus Concurrent Thinking-Out-Loud in Usability Testing*. Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1127-1131.
14. Pessin, J. (1933). *The Comparative Effects of Social and Mechanical Stimulation on Memorizing*. American Journal of Psychology, 45, 263-270.
15. Robinson-Staveley, K. & Cooper, J. (1990). *Mere Presence, Gender, and Reactions to Computers: Studying human-computer interaction in the social context*. Journal of Experimental Social Psychology, 26, 168-183.
16. Rubin, Jeffrey (1994). Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. New York NY: John Wiley & Sons, Inc.
17. Smith, M.J., Carayon, P., Sanders, K.J., Lim, S.Y. & LeGrande, D. (1992). *Employee stress and Health Complaints in Jobs with and without Electronic Performance Monitoring*. Applied Ergonomics 23, 17-28.
18. Triplett, N. (1898). *The Dynamogenic factors in pacemaking and competition*. American Journal of Psychology, 9, 507-533.
19. Wright, R.B., & Converse S.A. (1992). Method Bias and Concurrent Verbal Protocol in Software Usability Testing. Proceedings of the Human Factors Society 36th Annual Meeting, 1220-1224.
20. Zajonc, Robert B. (1965). *Social Facilitation*. Science 149, 269-275.
21. Zajonc, R.B., Heingartner, A., & Herman, E.M. (1969). *Social Enhancement and Impairment of Performance in the Cockroach.* Journal of Personality and Social Psychology, 13, 83-92.