A Multimethod Usability Approach to Evaluating E-Learning Applications – A Case Study

Marcel Vervenne, Mark Bormann, Michael Kerres

Chair of Educational Media and Knowledge Management
University of Duisburg-Essen

The paper describes a usability approach to evaluating E-Learning applications. In the context of an evaluation study of a computer based training on "leadership in public administration" accomplished for the ministry of the Interior of North-Rhine-Westphalia (Germany), a multimethod approach has been developed that combines traditional methods of empirical research, for example questionnaire and guided interview, with approaches from usability engineering, like eyetracking and annotation of multimodal behaviour. Although this approach is quite complex it reveals to be very helpful in improving the quality of an e-learning application. The paper points out that with this approach we are able to identify problems in using the software that would not have been recognized with a less complex methodology relying on a limited set of data. Some exemplary results form our usability setup demonstrates the potential of this procedure for formative evaluation.

1. Introduction

In 2005, the German federal state of North-Rhine-Westphalia (NRW) started a state-wide training measure for introducing new approaches of human resource development and to train leadership in public administration. Part of this measure is a computer based training program that is integrated in a blended learning arrangement. The topic of the CBT program was conducting dialogues and making agreements between staff members and executives.[1] The software was developed in two (only slightly different) versions to be used by either members of staff or executives. The software is conceptualized to address more than 200.000 employees in public administration in NRW. Since the large scale of the measure, issues of usability were given a high priority in the implementation process since even smaller usability problems might risk the success and the efficiency of the complete measure. It was very clear that usability problems would have impeded the success of the measure altogether. The formative evaluation of this ambitious project included testing and optimizing a prototype of the CBT, as well as investigating acceptance and learning-effects of the measure. Methodically this was implemented by a number of investigation measures as described in the following section.

[1] in German: „Mitarbeitergespräche führen – Ziele vereinbaren"

2. Method

Due to the complexity of the blended learning measure, a research design was developed within the framework of this study, which should illuminate the investigation object from a multidimensional perspective (cf. Voss et. al. 2006). The following figure outlines the four modules of the research design and the respective instruments:

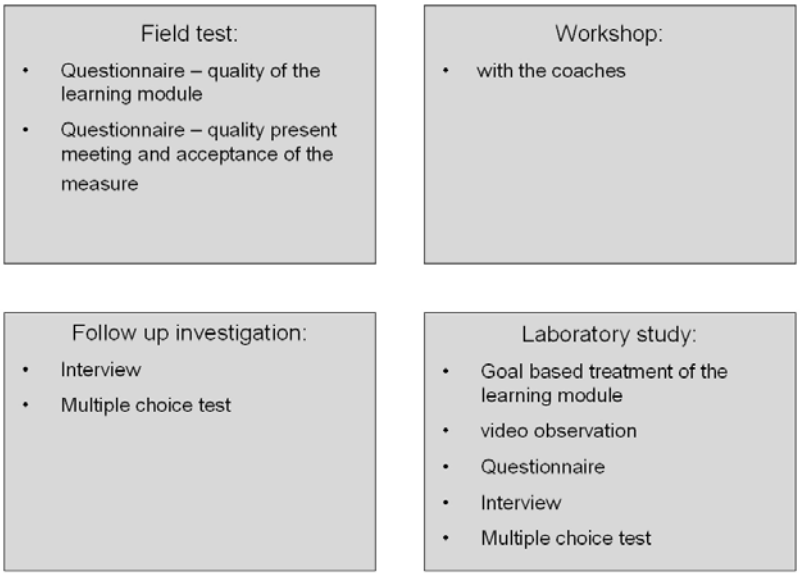

Figure 1: *the 4 Modules used within the study.*

In the context of the field study, three central questions were analysed: the usability of the CBT, the acceptance of the measure as a whole and the quality of the face-to-face (FTF) training. In addition to the field study, a one-day workshop with the coaches of the face-to-face training was accomplished. This workshop mainly focused on the question to what extent the previous use of the CBT really did prepare the learners for the FTF training adequately: Did the FTF benefit from the prior training with CBT? A follow-up interview with learners examined learning effects and addressed the question if the learners were able to apply the learning contents to their professional work life.
Apart from these three modules, a laboratory study was conducted to optimize the usability of the CBT in more detail. In this laboratory setting, learners were video-taped while working on a part of the CBT. Furthermore, the setup included eye-tracking. Using the think aloud technique with cooperative evaluation, all participants were asked for a continuous verbalization of usability issues occurring during the one hour test session. A questionnaire and a guided interview concluded the session. In total, 23 executives from different institutions took part in the laboratory study. All of them were novices with respect to the CBT. The following section describes the components of the setup in more detail.

2.1 The questionnaire

With respect to the content of the CBT, the questionnaire developed for the laboratory study covered a global evaluation of the CBT on the basis of a semantic differential, an estimate of the operability of the system as well as an evaluation of the relevance of the learning units. The questionnaire was filled out by the subjects immediately after the one hour session with the CBT program.

2.2 Think aloud and cooperative evaluation

"Think aloud is a form of observation where the user is asked to talk trough what he is doing as he is being observed; for example, describing what he believes is happening, why he takes an action, what he is trying to do." (cf. Dix et. al. 2004 P. 343)
Beside this classical form of think aloud, cooperative evaluation has been used here. By this variation, the interaction between the investigator and the subject is intensified to incorporate the person more strongly into the evaluation process. Verbalization was integrated with behaviour observation of the subject and analysed according to the usability principles defined in DIN EN ISO 9241-10. Along with these classical approaches in usability engineering, a Tobii 1750 eyetracking system was used to record the visual activity of the user during the study.

2.3 Eyetracking

The most basic question regarding the application of eyetracking methods in usability studies refers to the connection between visual reception behaviour and the correspondence of cognitive processes. One central goal, for example, is to infer cognitive processes like attention, stimulus complexity or data processing by means of measurable eyetracking parameters like the fixation spot, fixation time and -frequencies or the gaze path. Implicitly, such approaches take two basic - simplifying - methodological assumptions (cf. Just & Carpenter, 1980; Schroiff, 1986):

> *1. The visual attention focus of the user lies on the object, which is currently the object of cognitive processing (eye-mind assumption).*
>
> *2. Fixation time corresponds to the duration of the cognitive processing (immediacy assumption).*

Although there is empirical evidence for the validity of these assumptions, (cf. Just & Carpenter, 1980) their scope is still under critical discussion (for example Rayner & Sereno, 1994; Roetting, 2001). Nevertheless, eyetracking is regarded as a sufficiently valid method to generate objective indicators for a better understanding of information recognition and processing. So it is a useful method to generate recommendations for the optimization of systems – such as e-learning software - within the framework of formative evaluation.

2.4 Interview

After working on the CBT and filling out the questionnaire a guided interview was accomplished with the subjects. Here, verbal statements the subjects made while working on the program served as a basis for discussion. Very individual problems could be addressed, which could not be dealt with by means of the questionnaire alone. Beyond that, it was possible to talk more exactly about suggestions for improvements or modifications with the participants. These results directed the route for the inspection of the other sources.

3. Results

This section shows exemplarily the benefit of the combination of the selected methods by describing two critical usability issues.

3.1 Example 1

The analysis of the gaze path, also known as *scanpath* or *searchpath analysis*, currently forms the central approach to the analysis of eyetracking data within application oriented research (Bente, 2004). Fig. 2 shows an example of such an analysis for two subjects looking at the introduction of the learning software. The subject's main focus of attention is represented in terms of the rank place order of the visual contact with the corresponding area. Saccadic movements between the single fixation spots are represented as well. Additionally, the size of the respective fixation spot carries information concerning the fixation time in that manner, that bigger fixation spots represent longer fixation times.

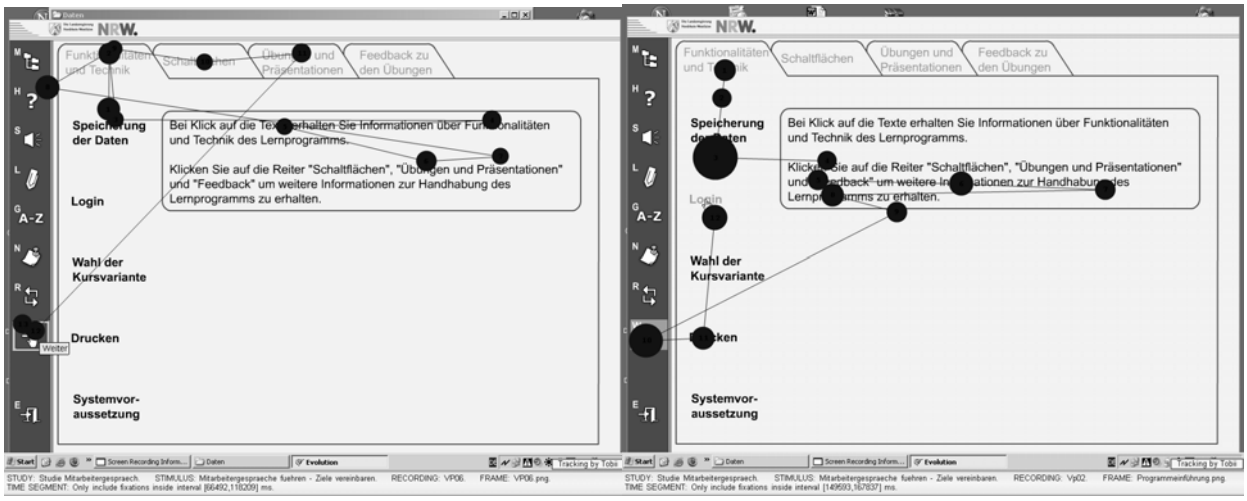

Figure 2: *gaze paths of two learners*

The navigational structure of the CBT assumes that the user takes a certain path through the application. First, the user is supposed to obtain basic information about to the software using the tabs that describe the application before s/he continues to gather further information about the topic of the program by using the button „next". In fact, the analysis of the gaze path reveals, however, instructive deviations from this intended reception. The initial fixation of most learners - as intended by the program – is in the upper navigation area. Also in line with the instructional concept, the content area is observed intensively by most users, which explains and invites them to the further use of the upper navigation elements. As the gaze path analysis shows impressively, a strong attention shift in direction of the main navigation menu at the left edge of the screen occurs shortly after the start of the application (visualized as strong saccadic movements in figure 2). The reason for this visual behaviour is the strong affordance of the button „next" which begins to flash approximately 10 seconds after starting the program. This misleads many users to a false association between the flashing of the button „next" and the nearby „print" button. This usability problem was verbalized only by a single subject on the think aloud technique ("Oh, the "print" button flashes, so I better do nothing"). However, a re-analysis of the eyetracking data – cf. figure 2 – eventually showed, that this usability problem by no means is an individual case.

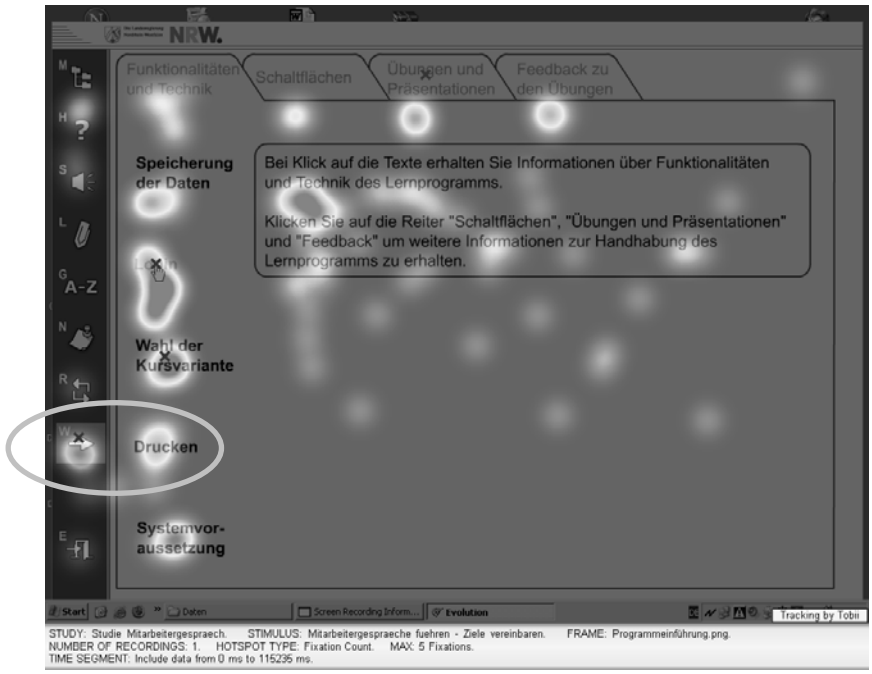

Figure 3: *Aggregated fixation frequencies (n=5 subjects) for the consideration of the CBT introduction. Zones of high visual attention are brightened. The areas "next" and "print" - which were wrongly associated - are also highlighted.*

This pattern of corresponding fixation frequencies for the spatially neighbouring areas „next" and „print", was observed by 5 users in total, even those who did not verbalize any mental association between these two functional areas within the framework of the think aloud method. There are several indicators from the gaze data that give proof to the hypotheses that this interpretation does (at least

latently) exist, e.g. the proximity of rank place order of first visual contacts with the respective areas. By pointing out user's latent cognitive processes these gaze data clearly identify essentially weak points of the CBT.

3.2 Example 2

The questionnaire presented after the learners worked on the CBT contains several items about the operability of the CBT. Approximately 82 percent of the users felt that the exercises the CBT provides are simple to handle (”the operability of the learning tasks and exercises is simple”), thus conveying an overall very good impression of the usability of the program. But in combination with the other data sources, it can be illustrated very well, that the questionnaire data alone would have been misleading since these verbal report data does not correspond with other results of the laboratory study. Actually, there were substantial problems while working on several learning tasks.

One of these tasks – and the problems associated with them – will be illustrated and discussed here. Within some of the learning units the user has to evaluate statements as being true or false. These exercises usually function as a test for evaluating the learning progress. The learner has to agree or to reject presented statements. After an option is selected the learner gets the feedback, if the answer was false or correct, shown in the lower section of the screen. This comment must always be confirmed by a single mouse-click into this field to enable the user to continue with the exercise. This procedure is clearly described in the program introduction. Nevertheless, *all* subjects actually tried to continue immediately with the exercise and were irritated by the fact that they could not make further inputs. One has to consider, that nearly all of them had read the necessary information how to deal with these tasks in the introduction before.

This phenomenon could clearly be identified on the basis of the eye tracking data. Above all, apart from the inspection of the behaviour the think aloud comments of the subjects emphasize the amount of irritation quite vividly. The problem of dealing with the exercises was also addressed by the majority of the learners in the interview.

This difficulty of handling the exercises showed up impressively in the observed behaviour of one learner who reads out the information text aloud (“If you receive a comment within an exercise you will have to click into the feedback text, in order to be able to continue with the exercise.”), but was not able to remember the information approximately fifteen minutes later while directly working on the particular task. This quite illuminating case - exposed by the method of think aloud – actually was the trigger to examine more exactly whether this was an individual case or not. The analysis of the corresponding eyetracking data of other learners then showed, that this problem was found in the patterns of *all* learners. The following figure of a gaze paths shows, on the left side, the perception of the tutorial, and on the right side, the failure to handle the corresponding task by the same subject.

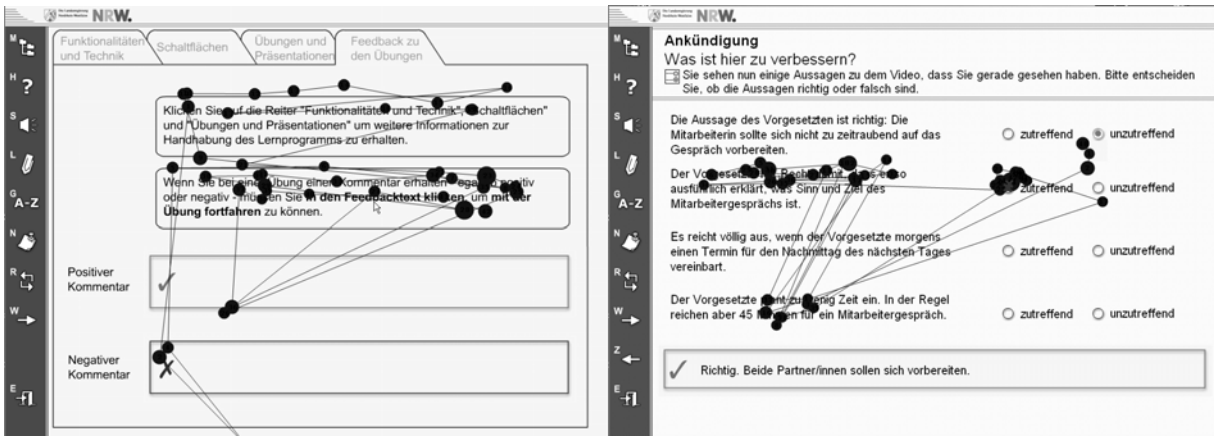

Figure 4: *2 exemplary gaze path of one subject working on an exercise of the CBT.*

4. Discussion

The case outlined above show how usability issues can be identified by a combination of verbal data and gaze movement analysis. With traditional usability inspection methods (cf. Nielsen, 1994) alone, the usability problems presented above might have been identified. But typically, they would have been classified as individual cases. With our multimethod approach usability problems can be discovered even if subjects are not able to or refuse to verbalize possible irritations within the think-aloud procedure. Here, the supplementary relationship of the different methodical approaches becomes evident. Verbal data from the think-aloud procedure, the questionnaire and the interview indicate usability issues that can be better interpreted in the light of objective data derived from gaze analysis.
From the case presented here, recommendations for a general test arrangement for formative evaluation can be derived. Verbal data from questionnaires, interviews and think-aloud procedures supply subjective references, which should be related to objective data from eyetracking. Thus, problems which would not be revealed by means of a questionnaire can then be recognized and, moreover, their consequences can be identified. For this reason, the different methods can be integrated into a cycle, where subjective data build the basis for the following inspection of the objective data. The inspection of the objective data alone, on the other hand, would not have revealed the usability problems easily. Therefore, the verbal data functioned as a pointer to inspecting possible usability issues in the objective data sources. The described laboratory setup appears rather complex (and expensive) on first impression. But in the context of formative evaluation the added investment pays off easily. With few participants in such a laboratory study very fundamental usability problems can be identified very fast. Therefore, we expect that in the future such a multimethod approach can be established as a routine measure in formative evaluation of software development.

5. References

[1] BENTE, G., Erfassung und Analyse des Blickverhaltens, In Mangold, P., Vorderer, P. & Bente, G. (Hrsg.), Lehrbuch der Medienpsychologie, 296-324, Hogrefe, Göttingen 2004

[2] DIX, A., FINLAY, J., ABOWD, G. D., BEALE, R., Human-Computer Interaction, Pearson Education Ltd., Essex 2004

[3] DUCHOWSKI, A. T., Eye Tracking Methodology: Theory and Practice, Springer, London 2003

[4] GRONER, R., D´YDEWALLE, G., PARHAM, R., From Eye to Mind: Information Acquisition in Perception, Search, and Reading, Elsevier Science Publishers, Amsterdam 1990

[5] NIELSEN, J., MACK, R.L., Usability Inspection Methods, John Wiley & Sons, New York 1994

[6] RAYNER, K, SERENO, S. C., Eye movements in reading: Psycholinguistic Studies, In Gernsbacher, Morton Ann (Ed); et-al. Handbook of psycholinguistics, 57 – 81, USA: Academic Press Inc., San Diego 1994

[7] RÖTTING, M., Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen, Shaker Verlag, Aachen 2001

[8] VOß, B., BORMANN, M., VERVENNE, M., KERRES, M., Begleituntersuchung zu „Mitarbeitergespräche führen – Ziele vereinbaren in der Verwaltung NRW [Online], available at: http://mediendidaktik.uni-duisburg-essen.de/files/blendedlearning-nrw.pdf , 2006