

Remote Usability Tests – An Extension of the Usability Toolbox for Online-Shops

Tim Bosenick¹, Steffen Kehr², Martina Kühn¹, and Stephan Nufer³

¹ SirValUse Consulting GmbH, Schloßstraße 8g, 22041 Hamburg, Germany
bosenick@sirvaluse.de

² Otto GmbH & Co. KG, Neue Medien / E-Commerce, Wandsbeker Str. 3-7,
22172 Hamburg, Germany
steffen.kehr@otto.de

³ Studiengang Digitale Medien, Fachhochschule Kaiserslautern, Meyerbeerstrasse 6,
22145 Hamburg, Germany
mail@stephan-nufer.de

Abstract. Lab Usability Tests show a series of inherent shortcomings that are attributable in essence to the artificial lab situation. This article informs about the reasons for developing a specific Remote Testing approach and describes how this measure helps to avoid such deficits. Subsequently, we will introduce the approach as well as two evaluation studies that assess the result quality of a Remote and a Lab Test within the context of online shopping.

Keywords: Usability Test, Remote Test, method comparison, online shop.

1 Introduction

Testing interactive web applications assisted by the current respondents in the Usability Lab probably presents one of the most popular methods of the Usability toolbox. Tests of this sort provide valuable qualitative results based on user observation and inquiries. User interface designers, for instance, are able to translate these results smoothly into an optimisation of the application. In the meantime, User Tests have become well-accepted by all persons involved and form an increasingly steady component of developing interactive applications, especially for their high ostensible validity.

In general User Tests take place in a classical Usability Lab equipped with a lab plus observation room and comprise the production of a video tape recording. A test head observes, interviews, and instructs the participants in solving the particular type of problem. Apart from many strengths, however, this Usability method also entails some restrictions that make it seem less suitable for certain application purposes.

The artificial lab situation in particular may influence the respondents' behaviour considerably thus endangering the translation of the test results into the real context of use. Moreover, the circle of available participants often proves strongly limited with respect to their number and composition (background of use, level of experience, socio-demography). Recruitment and test conduction are characterised by tight boundaries installed by necessary expenditures, or rather, by the financing of their

costs. For the past year SirValUse and Otto have already been trying to circumvent these limitations of Lab Tests by the joint development of a tool for asynchronous Remote Usability Tests. During an asynchronous Remote Usability Test the respondents are testing a web application from their own PC or from their workplace. They are totally independent in their moves on the test object without any time limit or content restriction. The Remote System records all mouse movements, clicks, and entries as well as screenshots of all pages visited by the participants. The respondents are requested to report any special occurrences (so-called 'critical incidents'), such as obstacles or missing functions, while they are surfing into an entry mask. The following text will describe the developed Remote Test System, the exemplary course of a Remote Test, and the results of the most recent comparison of our Lab Test and Remote Usability Test methods.

2 The Path to Gaining Insight – User Tests in the Course of the Project

During the development of an interactive application, or rather, beyond the life cycles of a product, the persons involved in the project are confronted with very distinct user- and use-related questions that are to be answered based on User Tests. The demands on User Tests are thus as diverse as their implementation and arrangement.

2.1 Rough and Fine Concepts as Subjects of Development-Oriented User Tests

Usability Tests of interactive rough concepts focus primarily on the question: "Are we on the right track?" Early conceptual ideas are examined rapidly without a drain on resources. Nonetheless, the tests ought to provide a stimulus for the project team's ability to cope with problems.

In case the development has already reached the final stages with tests of the fine concept, the pivotal question is: "Are we really finished with it?" At this point the User Test must be fit to assess the Usability of the entire application to enable the development team to decide for or against the launch without a doubt.

Development-oriented User Tests can be conducted ideally either by means of paper prototypes, by employing more or less interactive prototypes in the Usability Lab, or next door to the developers.

2.2 Status-Oriented User Tests

Proving whether an application reaches the desired business- and customer-related project goals during day operation, requires another type of User Test that asks: "Can the desired effects indeed be measured in the actual applicants?," " To which degree could the application be improved?" In sum, we have to provide convincing quantitative evidence that the aspired business-relevant effects have in fact ensued during the actual operation.

In the past we have conducted slightly modified Lab Tests, differing in certain characteristics from development-oriented Lab Tests, in order to answer the above questions as regards online shop purchases.

For this kind of User Test we have recruited respondents with actual purchase intentions of their own and asked them to conduct their planned online purchases under observation in a Usability Lab. The participants did not get tasks and remained absolutely free and uninfluenced by the test head who left the test lab after instructing them. In contrast to the classical User Test the respondents were to decide for themselves how to proceed, how much time they needed, and if they really bought anything at all. In coherence with the underlying logic they invested their own money for their purchases.

The above modifications and further adaptations were intended to equip User Tests for simulating a realistic buying situation as a means to assess the Usability status of the online shop. This approach, however, came up against limiting factors repeatedly:

- The circle of respondents for these tests is strongly limited in various respects. For financial reasons, for example, the sample is confined to about 30 participants which is too small for differentiating quantitative assessments.
- The tests are conducted in one location only.
- Purchase processes that can actually be observed in the lab are quite restricted. For instance, none of the processes can be verified after completion of the order. Neither is the delivery status being followed, nor are returned sales commissioned, nor is the customer account being checked for the correct entry of the sums. For that reason, in a Lab Test the test head can only retrace a limited selection of all destinations a customer may actually visit in an online shop.
- The lab situation remains an artificial construct: the time and duration of the shopping session, the environment, the technical equipment, and of course the presence of the test head all represent very artificial components of the shopping situation. These conditions differ significantly from those under which customers of the shop would decide about their purchases in reality.
- Despite the above mentioned recruitment efforts, actual and independent buying decisions cannot always be observed in a studio. Sometimes, further consultation with the partner may seem necessary, the respondent may not have fully decided yet, or the incentive may appear more attractive than the order itself.

These limitations, in connection with our desire to find a test method for online shops that is more valid in ecological respect, have led to the development of the asynchronous Remote Usability Test method that will be illustrated in the following.

3 Asynchronous Remote Usability Tests

A Remote Test is a Usability Test that enables the participants to either use an application independently or solve certain problems, not in a lab, but at their home PC or from their work place. This type of test is called synchronous if a Usability expert, who may even be connected with the respondents by phone, observes the use situation simultaneously. A Remote Test is asynchronous, if the expert observation is subsequent to the actual session.

In comparison with the synchronous approach the asynchronous test has the advantage of allowing that a great number of respondents take part in the test and of providing very focused analyses. To its disadvantage it does not permit any additional,

immediate questioning, for instance via a parallel phone connection. Any language channel would disturb what can be considered the greatest advantage of the asynchronous approach: the authenticity of the setting.

The system developed for the evaluation studies distinguishes itself for the following technical characteristics:

- The users' requests are first transferred to the web site via a proxy server. The server then hands the web site back to the respondent while simultaneously recording the users' screenshots, mouse movements, and entries at that moment.
- Hence, installations or interventions on the web server and the participants' PC can be avoided provided that Javascript is activated in the respondents' browser software.
- The test object must be accessible online.

3.1 A Remote Test Session Then Proceeds as Follows

- The participants are recruited online by means of a layer, a pop-up on the home page, and/or on various locations within the web site.
- After the general introductory text passage and the declaration of data privacy protection relevant screening questions (for example socio-demography, internet use, visit frequency, customer status, etc.) as well as questions for the general pre-exploration (mostly: reason for visiting) are established.
- It follows an explanation of the session's planned process: the respondents are asked to move about on the page according to their initial intentions. Moreover, they are familiarised with special occurrences ('critical incidents') and how to report them while surfing.
- A click upon a link activates the re-routing onto the Remote System. The web site is depicted in a frame that covers the biggest part of the screen. Another, smaller frame at the bottom rim of the screen offers buttons for reporting 'critical incidents.'
- The feedback questionnaire appears as a pop-up that gives room for reporting the 'critical incident,' its 'pre-history,' plus maybe a scaled evaluation. In the course of a session any number of 'critical incidents' can be reported.
- If the visit is completed, a button in the bottom frame leads to a post-inquiry about the visit's success and asks for a summarised evaluation of the web site.

Prior to the analysis the different data sources (screenshots plus mouse movements, reported 'critical incidents,' and further enquiry data) are brought together for each individual respondent and cumulatively for all participants. The analysis can be applied specifically to the following data:

- the cumulated log files,
- 'critical incidents' in connection with the corresponding screen shots, mouse movements and entries,
- the participants' click routes with negative visit success, and
- further interview data.

The analysis includes not only subjective data ('critical incidents,' enquiry data), but also observations and compilations (ex-post analysis of click routes and log files) This is of pivotal importance especially as regards the compilation of Usability problems.

4 Comparison of Lab Test and Remote Usability Test Results

The advantages of the Remote Usability Test in general and as applied to online shopping in particular seem to be perfectly obvious:

- Online recruitment permits a simple and well-controllable access to a wide spectrum of respondents.
- The observation comprises very different visit destinations, covering every phase of the shopping process. Accordingly, the scale of identifiable Usability problems is quite diversified.
- The casual testing in familiar environments without set tasks and a test head creates a high ecological validity.
- By focusing on the reported 'critical incidents' and those test sessions that have been evaluated as 'unsuccessful' the test sessions can be evaluated very efficiently.

On the other hand fundamental questions concerning the method arise:

- "Are users able to detect and report their own use problems?"
- "How efficient are Remote Usability Tests in revealing Usability problems?"

We conducted 2 evaluation studies to answer these questions. For both studies a Remote and Lab Test for an online shop were conducted simultaneously and their results compared.

The more recent, second study was conducted following the procedure described above. In this context 111 User were tested compared with 30 respondents in the lab. For the initial evaluation study 30 customers of the shop were recruited offline (by post) and asked to use the online shop through the Remote System for a period of 6 weeks. In this case the Lab Test included 15 participants.

4.1 "Are Respondents Able to Detect and Communicate Use Problems on Their Own?"

An asynchronous Remote Usability Test relies substantially on the participants' reports of high quality 'critical incidents.' The two evaluation studies produced the following results (the first number represents the results of the recent, second study, the second number the results of the initial study):

- Altogether 48 percent (91 percent) of all respondents recognised and reported 'critical incidents' which resulted in identifying Usability problems.
- 87 percent of these reports allowed us to comprehend which incidents were described and why. We attested them a high quality due to their comprehensible wording, their being voluntary, and the fact that they could be divided unequivocally into positive and negative reports; furthermore they withstood the inspection of a Usability expert.

The low rate of respondents who report 'critical incidents' at all increases when positive incidents are also taken into account; the participation rate then reaches 69%. By means of an adapted incentive scheme this rate could maybe further increased. It is another advantage of the Remote Test that without any incidents the participants do not cause any additional expenditures apart from the incentive.

In addition to 'critical incidents' there is another observation component that contributes valuable insights: the later analysis of the click routes that considers the reasons for visiting and the visit success (especially failure).

4.2 “Is a Remote Usability Test the More Efficient Method in Unearthing Usability Problems?”

The results show that more Usability problems are detected during a Remote Test than during a Lab Test: 61 (28) problems in the Remote Test opposite 49 (25) problems in the Lab Test. In essence this derives from the greater number of participants in the Remote Test and their more differentiated reasons for visiting.

It is remarkable that the problems revealed by the Remote Test do not concur with those of the Lab Test in many respects.

In the real and very diversified context of use the Remote Usability Test finds Usability problems which the respondents consider worthy of report. The Lab Test, by contrast, detects Usability problems that are generated mostly by the observation of participants who find themselves in a (more or less) artificial situation. These very distinct ways of proceeding obviously effect the differing result quantities of Usability problems. (See Molich¹).

With respect to the arising expenditures as related to the identified Usability problems and the participants involved, the Remote Usability Test proves much more efficient than the Lab Test. By means of a partially automated summary and data evaluation this advantage can be consolidated further.

As regards the investigative gain the Remote Usability Test is inferior to the Lab Test. Half of the respondents do not produce any or just a minor investigative gain (for instance, the evaluation of the reason for visiting and the success thereof). In the lab, however, only 29 (52) percent of the participants are unable to contribute any essential insights. Nevertheless, this effect is compensated by the Remote Usability Test's cost-effectiveness per respondent. Overall each Euro invested in a Remote Usability Test finds more Usability problems than it would in a Lab Test.

Concerning the speed of the operation the Remote Usability Test has the potential to unearth Usability problems within a short time - faster than the Lab Test. The rapid online recruitment and the test procedure that employs several participants in parallel, shorten the period of testing significantly.

5 Prospect

Even after the conduction of 2 evaluation studies many questions still remain unanswered and need further clarification:

¹ Molich, R.: CUE Studies Homepage, available at: <http://www.dialogdesign.dk/cue.html> (1998 – 2006) (05.22.2006).

- “How helpful are problems found by both methods as regards the business case, or rather, the customer's investigative interest?”
- “Which method provides better answers to which type of problem?” Current studies are indicating that especially the evaluation of web site contents becomes more valid in a Remote Test. So, “What characterises the problems that are detected primarily by a Remote, or rather, by a Lab Test?”
- “Which problems ask for a combination of both, a Remote and a Lab Test?,” “Which case number corresponds with this?”
- “Which is the optimum case number for a Remote Test?” “When will a decreased investigative gain start to kick in?”
- “How to increase the number of reported 'critical incidents' per respondent without provoking any irrelevant reports?”
- “Which role could the ex-post examination of the respondents' click routes play in relation to the reason for visiting and the visit's success?,” “Can the number of problems detected by both methods thus be increased?”

Accordingly, we are looking forward to further evaluation studies.