

# Usability Testing of e-Learning in Practice: First Experiences and Lessons Learned

Stephan Weibelzahl  
National College of Ireland  
Mayor Street  
Dublin 1, Ireland  
+353 1 4498 579  
sweibelzahl@ncirl.ie

Abigail Reynolds  
National College of Ireland  
Mayor Street  
Dublin 1, Ireland  
+353 1 4498 664  
areynolds@ncirl.ie

## ABSTRACT

As part of National College of Ireland, we have recently opened the National e-Learning Laboratory (NELL), a research facility specialising in usability testing. Since its opening in December 2008, researchers in NELL have been involved in a large number of usability projects with a variety of different clients. As testing has developed, we have been investigating various methodologies of how to evaluate products being tested and how to best describe user experience.

The aim of this short paper is to present our first experiences of usability testing; to describe our research methodology and to discuss lessons learned. We look at some common themes that have emerged from the usability research projects; such as the extent to which user expectations are guided by previous experience and how, when the system did not conform to prior expectation, this had a considerable negative impact on ease of use and satisfaction. We also consider some lessons we have learned as researchers while undertaking the testing, such as the usefulness of user personas and scenarios when designing tests and the importance of making user tasks non prescriptive.

We conclude by looking at the issues concerning usability testing specifically related to e-learning, where surprisingly a coherent and widely used usability framework is not yet available. We consider what is necessary for the future development of such a framework.

## General Terms

Experimentation, Human Factors

## Keywords

Usability testing, e-Learning framework, usability laboratory, eye-tracking, user experience

## 1. INTRODUCTION

In the National e-learning laboratory (NELL) investigators engage in applied research directed at user-oriented testing to support good design for e-learning, e-commerce and Internet based applications. Researchers can evaluate the students' learning experience in terms of the quality of engagement, efficiency, effectiveness and usability.

To date, researchers in NELL have been involved in a variety of different usability research projects with organisations both large and small. We have worked with a number of SME's on different projects ranging from usability testing of on-line maths and accountancy support programmes through to a research project analysing what people look at when presented with the results of a Google Search page. Many of the research projects involving SMEs were funded by the Innovation Voucher Initiative run by Enterprise Ireland, which encourages research collaboration between small firms and third level providers.

Larger institutions such as NALA, the National Adult Literacy agency, also recognise benefits of usability testing and have just completed a usability research project with NELL, looking at the usability attributes of NALA's new on-line literacy tool, [www.writeon.ie](http://www.writeon.ie), a website to support service users improve their reading, writing and number skills.

As testing has developed, we have been investigating various methodologies of how to evaluate the products being tested and how to best describe and report on the user experience. We are particularly interested in understanding more about e-learning and as such been investigating ways to classify and test usability aspects that are specific to e-learning applications.

## 2. USABILITY TESTING IN NELL

### 2.1 Research Design

At the start of any usability project, we first worked with clients to identify a number of critical tasks that regular users of their systems would need to perform. We aimed to make sure that test users would be representative of the target group and where necessary match a given persona. Either the client would provide us with test users, or for most part, we recruited test users ourselves. For the majority of these studies, we recruited five testers per project. In all our usability projects test users were defined as 'new users' of the system as opposed to returning users/learners.

Once tasks were defined, we gave participants instructions on what we would like them to do during the test. They were presented with a scenario and asked to complete a set of pre-defined tasks. Although we did ask users to complete a number of activities, we sought to keep tasks general and not too prescriptive as we wanted to emulate a scenario were users were new to a website and would need to explore various avenues before finding the solution. If the participants encountered a problem during

testing or could not find the answer immediately, we encouraged them to try alternative routes before seeking help.

## 2.2 Data Collection

Using a combination of video and audio recording, screen-capture, keyboard & mouse logging and eye tracking equipment, users' behaviour was observed and recorded. Audio and visual data were collected using Observer XT technology, mouse and keyboard events were logged in a covert manner using uLog and eye-tracking data was produced using SMI Experiment Suite software.

Eye-tracking is an interesting way of analysing user behaviour. It allows to measure and record what people actually look at rather than having to rely on think-aloud protocols or post-hoc interviews. Information like gaze duration, number of fixations and sequence of exploration is not easily available in any other way and can provide unique insights into what users are actually doing when looking at search results.

Prior to the test in order to collect data on demographics and users prior experience of the system, we asked users to complete an on-line survey. At the end of the test, we conducted a post-test interview with each user, asking them to run through various activities of the test and focusing on any problem areas identified during the session.

In order to start evaluating user behaviour, we first looked at the standard usability attributes such as those defined in the current 9241-11 ISO Standards (1998) which refer to Efficiency, Effectiveness, and Satisfaction:

- effectiveness (the ability of users to complete tasks using the system, and the quality of the output of those tasks),
- efficiency (the level of resource consumed in performing tasks),
- satisfaction (users' subjective reactions to using the system).

However, using these criteria alone we found it difficult to translate our findings into meaningful recommendations for the client. Thus, we broadened our evaluation criteria to include more subjective measures of analysis such as fun factor, levels of engagement etc., attributes similar to those mentioned in the 'experience' side of Kirakowski and McNamara model of interaction [1]. When testing e-learning products, we were unable to find any widely used evaluation criteria that could be specifically associated with e-learning. We therefore sought to develop and test our own set of attributes that we considered appropriate for e-learning such as progression, motivation etc. We

included these questions in our post-test questionnaire.

We then examined these issues in more depth in the post-test interview where we requested users to tell us about their experience with the system in general and asked them to comment on aspects such as ease of use, satisfaction, design, motivations etc. It was only when we combined all of these factors mentioned above in our analysis, did a more meaningful picture emerge from the data.

## 2.3 Analysis

When undertaking analysis, we looked at these data from several different sources in parallel. User behaviour was recorded and analyzed using Observer XT software.

This software allowed us to develop a coding system for each project so we could quantify and classify events being observed. We observed whether users were on task or not on task and recorded user errors, hesitation, facial expressions and gesture.

Based on these observations we attributed a 'success rate' to each of the given tasks. Ratings range from "task completed efficiently" and "task completed with hesitation" to "required intervention" being the least successful rating. The results of this analysis was then cross-compared to feedback from interviews and post-questionnaire.

A summary of all of these data can be represented in an observation table similar to Table 1 for each user. This format integrates the different data sources and provides a good overview of each user's experience.

This in-depth analysis and triangulation requires substantial effort but helps paint a fuller and more accurate picture of user experience. Many times, we would observe users having difficulty with a particular task, but when asked to rate task or comment on in the interview, the user would report they did not have any difficulties. At the same time, the post test interview was extremely valuable, as it added depth to data, giving the participants an opportunity to explain their actions during the testing. Furthermore, commentary from the post-test interviews often gave us substantial additional data on areas of the system which we might not have initially thought about including in analysis, issues such as trust privacy etc.

For each project we produced reports detailing the user experience and highlighting any areas for improvement. In final reports, we first presented a section on each of the users' experience. We then looked at behaviour of all users across all tasks and looked at the usability of the system as a whole. The final section of the report discussed any trends which emerged from the data and highlighting any identified areas of difficulty, as well as areas of strength.

**Table 1: Example of an observation table integrating various data sources**

task	time	errors	facial expression	post-interview	comment	success
1	20	0	neutral	"That was easy."	efficient completion	1
2	92	1	confused	"I didn't know how to open the next lesson"	unable to progress, required intervention	5
3	...	...				

### 3. LESSONS LEARNED

#### 3.1 Themes that Emerged from the Data

To date we have completed nine different usability studies in the National e-Learning Laboratory at NCI. All these studies have been different in terms of the type of product/service being tested and the different users groups involved. Many of the resources were at different stages of development i.e., one was at a wire-frames stage of development, while others were either BETA version or had been up and running for several months.

As each project was context specific, the tasks varied according to the resource being tested and many of the trends that emerged from data were project specific. However, we did see some similar themes emerge around some of conventional usability criteria such as ease of use, efficiency, effectiveness and satisfaction:

For example, if the product did not comply with the users' expectations in terms of structure, they found it difficult to understand the concept quickly. Our testers tended to give up after a very short while. We found this to be the case especially with the so called 'Google Generation' e.g. the younger users.

We also saw how users' expectations are guided by previous experiences with other products and applications. For instance, users would expect that tool tips are available when hovering over buttons and components. If they found this not to be the case, they would get frustrated and annoyed. This is not a new finding and has been reported earlier in the literature [2] however, we were astonished by the reoccurrence of this finding.

We also got some interesting feedback regarding the Search functionality on websites. In one project in particular, a website which supplied information to educators working with deaf children, our testers expressed much frustration with how this function was displayed on the website. In this particular case, the search functionality was located in the middle of the page underneath a blog calendar. Our testers reported that their expectations would be that the Search function to be highly visible and at top of the page. Three of five testers failed to notice it, which is surprising considering it would have made it easier for them to complete their tasks. Once again, the system not being conform to prior expectations had negative impact on user satisfaction.

#### 3.2 The Art of Usability Testing

As we continue to conduct usability research and refine our methodology, we are interested in learning more about good practice in relation to usability testing. Since we have started research in this area, we ourselves have made a number of observations in regard to the good design and practicalities of conducting usability testing:

Given the limited number of participants we observed per project, we found that recruitment and selection of tasks are crucial to the success of the test. Participants obviously need to match the target group. The use of personas and scenarios can help to define and agree on required characteristics of users.

Observations should be as 'natural' as possible. The fact that participants are brought into the lab and that they know they are being observed might impose a bias in behaviour. Tasks should be kept general. Overly prescriptive tasks restrict users in what to do

and may result in behaviour artefacts that do not reflect real world behaviour.

Post-observation interviews were pivotal in gaining further understanding of user behaviour and facilitating the correct interpretation of the observations. In a number of cases, only the comparison of observed behaviour and feedback from interviews would reveal what was going on. For example, in one project a participant seemed to complete a task efficiently (i.e., navigated to correct area), but he later revealed that he had misunderstood the task and just happened to land in the right area. These interviews also gave us further insights into issues that were important to the user but which, we ourselves may have overlooked in the research design, issues such as those related to trust, privacy, security etc.

### 4. CONCLUSIONS

Over the course of the projects, the way we analyse and report observational data shaped up continuously. However, we are still looking to further improve our procedures. In particular, as a centre for technology enhanced learning, we are especially interested in understanding more about e-learning as well as about usability testing specifically related to e-learning. Although there have been some developments in producing a framework to evaluate e-learning applications (e.g., [3], [4]), to date a coherent and widely used framework that fully captures the peculiarities of these applications is not yet available.

Our aim is to look at developing a generic framework for usability testing of e-Learning products and resources. Such a framework would integrate usability criteria and user experience into a coherent way of data collection and analysis. It would guide researchers and practitioners in selecting appropriate tasks and criteria, yielding relevant and meaningful results. It would thus facilitate cross comparisons of e-Learning platforms, environments and courses.

From our experience to date, we have identified a number of areas that appear to be of particular concern for e-Learning applications such as progress and results, feedback on performance, as well as the visualisation of the learning path. These serve as a useful starting point in the development of an e-learning framework. As we continue to investigate these areas we would hope to join forces with other people in the scientific community to achieve the goals outlined above.

### 5. REFERENCES

- [1] McNamara, N. and Kirakowski, J. (2006). Functionality, usability, and user experience: three areas of concern. In *Interactions*, 13(6), 26-28
- [2] Jokinen, K. and Hurtig, T. (2006). User expectations and real experience on a multimodal interactive system. In: *INTERSPEECH-2006*, paper 1815-Tue2A3O.2.
- [3] Ardito, C., Costabile, F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V. 2006 An approach to usability evaluation of e-learning applications. *Univ. Access Inf. Soc.* 4(3), 270-283.

[4] Sullivan, J., Hall, R. H., Hilgers, M. G., Luna, Ronaldo., Buechler, Matthew R., and Lawrence, W. T. (2005). Iterative Usability Evaluation Methods Applied to Learning

Technology Development, In: World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), 2005.