Bridging the gap between literature and practice: usability testing

Karin den Bouwmeester[a] and Guido Stompff[b]

[a] *Usability engineer at Industrial Design Department Océ Technologies, Venlo, The Netherlands*
[b] *Product designer at Industrial Design Department Océ Technologies, Venlo, The Netherlands;
Teacher at Faculty of Industrial Design, Technical University, Eindhoven, The Netherlands*

Abstract
There is a large discrepancy between literature about usability testing, and usability testing in practice. Every designer or usability engineer knows for example that a test should be done with real end users. In practice all kinds of practical limitations (issues of confidentiality, lack of time, difficulty finding participants) often lead to testing with co-workers or students. Better this type of testing then no testing at all, right?

This article describes some of the difficulties of usability testing during complex product development. By analysing real life examples the limitations of practical testing are described. One of the Océ cases describes a usability test were co-workers were used in stead of actual users. This test led to "improvements" that later turned out to be "unusable" after a test with actual users. Can this situation be prevented? What difficulties can be overcome? When should the practitioner decide not to test? What can still be learned if we take the potential limitations into account?

The challenge is to create a form of usability testing that usable in practice without loosing it's usefulness.

1. Introduction

In the development of complex products such as printers, scanners and copiers usability aspects and ergonomical aspects need to be balanced with other requirements, within the timeframe a project must be realised. In practise this implies that there is a large discrepancy between literature and practice concerning usability testing when it comes to product development. The typical dynamics of product development impose constraints on the testing. The challenge therefore is: how to adapt usability testing to the nature of product development optimally, so that the usability of the end product will improve?

At Océ since some years usability research and testing is considered to be an important discipline within a development team. This paper is a case study concerning the experiences of Océ to integrate usability testing effectively into the product development.

2. About Océ

Océ is one of the world's leading suppliers of professional printing and document management systems. For offices, industry and the graphics market the company develops and manufactures systems for the production, distribution and management of documents. This includes printers, scanners and copiers, in colour and black and white, in small format and in wide format. Besides Océ supplies printing media and peripheral equipment but also document management software, innovative products in the

areas of system integration and outsourcing of document management activities. The company, which is commercially active in 80 countries, has built a world-wide reputation as an innovative business in both a commercial and a technological respect. [1]

3. The challenge of usability testing within product development

Complex products, such as DVD players, printers and medical equipment have in common that a wide range of knowledge areas is required to develop them, including interaction design and product design. In general this implies (large) multidisciplinary teams that need to find a balance between all sorts of conflicting requirements. The usability of a product surely is an important, but not the only requirement to be met. Consider the costs of a product, the safety to use it, the functionality it has or even the need to meet international laws and regulations.

We define, comparable to the definition [2] of ergonomists, usability testing as the discipline contributing to the design of products and systems, by means of evaluating them, in order to make them compatible with the needs, abilities and limitations of people. This underlines that usability testing (at least within product development) is not a goal on its own, but its intention is to influence the product development in such a way that the usability of the end product is maximised.

By reading the usability literature [e.g. 3] some do's and don'ts can be learned about how to organise a valid usability test, such as: a) test early; b) test with real users; c) test with sufficient users, etc. Although every usability engineer or ergonomist working for a company knows the above, they probably all recognise that in 'the real world' things are not that easy. Now what is it that makes it so hard to perform good usability testing inside the development of complex products?

In short: methodical usability testing requires ample time, functioning and stable prototypes with real users in an almost real life situation, doing well defined tasks. However: in a product development process these requirements only can be met at the end, when the project is almost finished and apart from some repairing no profound changes can be implemented anymore. Besides product development is in its nature a rather fuzzy process, where many choices influencing usability are made on the go, details changing constantly, some of which have considerable impact upon the final product. But not everything can be tested to evaluate whether or not it is an improvement. More companies acknowledge this problem and e.g. Microsoft uses [4] the RITE

approach for the development of software in which usability testing is in a project stage embedded within the design process. Changes are made as soon as a problem is identified and a solution is clear, sometimes even after observing one participant.

The challenge for usability testing within product development is therefore to find a good balance between the time and effort you put into it and the expected pay off, within the constraints imposed by the nature of development. This balance is fragile: too many concessions result into an invalid test, as we will show. Or when the test requires too much time and effort before test results are delivered, the decisions have been made already. In the next chapters some Océ cases are described in which this search for optimal balance between effort was sometimes successful, sometimes not so much.

4. Participants

One of the mayor difficulties of usability testing is finding enough and the right participants for your tests. But companies as Océ do not want to demonstrate too early to the world what they are working on, so there is an aspect of secrecy. Even more difficult there are often intellectual property matters involved (patents). Due to the nature of patent rules, a product innovation can not be shown before the procedure is done. Both quite different aspects make it very difficult to test with real users early enough, let alone in a real life situation.

Picture (1) *shows a test where we wanted to learn whether or not users could comfortably reach the upper output tray of a printer. The answer to this anthropometrical question is reasonably objective and we therefore tested with a range of Océ employees.*

Sometimes it seems necessary to compromise and search for alternatives, e.g. within your own organisation. Maybe some new employees of staff members who are not involved? Better to test with

these people then not to test at all, right? In some cases this is certainly true. For example if the test is not dependent on experience of the user e.g. when it concerns anthropometrical issues.

In the past we learned that sometimes it is dangerous not to use real users. For a high-end professional printer a new interaction concept was developed. On the local user interface it showed the 'Schedule', a matrix like representation in which the operator could find much information about how long the printer would keep on running and when intervention was needed by the operator, such as refilling or changing paper. Due to patent research for this new concept, it could not be shown yet. The design-team decided not to wait with a test, so a usability test was performed in which a number of Océ employees were asked to participate, most of them from R&D.

Picture 2. *The set up of the test performed firstly by Océ employees and later with professional operators. The screen shot (LUI) is the 'Schedule', a tool to be able to plan and program ahead.*

We thought the concept was a success, as it was understood easily and appreciated. Even more the test persons used it almost as a sort of game, trying to sort out how to do the tasks as effectively as was possible. The results indicated that additional information was needed to even plan activities better, resulting in some design alterations.

Some months later it was possible to test with real operators. Operators where invited and a comparable usability test was performed. The results from this test were completely different from the first however. The concept was hardly understood and the operators missed information they considered important, such as the ability to interrupt a job easily without loosing any information. The ability to program ahead (including all alterations since first test) was too difficult to understand. Working methods amongst the customers differed a lot, but they shared the need for at least feeling 'in control', more then the smart planning tool. The difference between the educated R&D people with a liking for abstract models and the down to earth operators doing the work every day was considerably disturbing test results. Much of the design effort between the two tests was more or less heading the wrong way.

It took considerable effort of the design-team to address these problems. Fortunately the testing was done on moment changes were still possible. The design team did not skip the 'Schedule' idea, being convinced that it was increasing the usability, but considerable effort was taken to make the 'Schedule' more easy and matching operator expectations. The third test we performed with the professional system recently showed that the operators appreciated the scheduler. The lesson to be learned is not that if you can not involve real users you should not test, but one should be careful how to interpreted the results of this test.

Picture 3. *The final interface with the 'Schedule'*

Another discussion about participants in usability tests that is written a lot about is the number of participants needed. The Nielsen's claim that "Five Users are Enough" [5] is well known, although this is thumb of rule is scientifically heavily disputed ever since then [6 and 7]. In product development this discussion seems not to be relevant. In the several years we performed usability testing we experienced that after observing 5, maybe 6 participants, the added value of more test persons no longer weighs up against the additional effort (finding the participants, performing and observing the tests, analysing the

data). Better to wrap it up and make sure the test results are implemented.

5. Prototyping

Often complex products are developed in modules and the final product is the sum of these modules. As an example: a printer has without doubt no value without relevant desktop software. These modules are often not developed within one team, but by a range of teams – sometimes literally oceans apart. The main goal of developing modular is to be able to develop and test without the other modules. This implies for usability testing that not until the last phases of a project there are integrated prototypes of a product / system to test – too late to influence the direction of the project anymore.

In the past we were convinced that the evaluation of products on its usability was only valid if there was at least a partly functioning prototype. You can not test a product when it is not reacting on your commands, including real prints and copies, sounds and errors. We had to wait until engineering prototypes before we could conduct tests or we had to invest deep into creating elaborate interface prototypes that were capable to communicate with prototypes in an earlier stage. Either way the testing concerned only the validation of the concept due to timing and investments made. It was virtually impossible to make any significant changes to the design anymore.

Picture 4. *Paper prototype in a wooden dummy interface. The 'screens' could be changed easily when a user had chosen something.*

During the last years we changed this policy and felt the need to test iterative and starting early in the development phase. There are several options to prototype and shown by means of a recently launched office printer / scanner / copier:
•   Paper prototypes [8] can be used to test whether or not users understand underlying concepts of interfaces, expectations and user needs. One of the

additional advantages of testing with sketch-like prototypes is that the participant understands that the product is still in the early design stage and feels more freely to point out negatives.
Dummies: dummy-products with prototyped interfaces. The added value is that for the test users the interface already has a high level look and feel including dynamics and is already on a (sort of) product, giving rich visual cues. One can really observe what people do to get their tasks done and the anthropometrical quality. The risk though of interpreting these tests is the lacking 'behaviour' of the printer. A printer makes certain noises (giving additional feedback), or it takes a certain time to perform an action (what can make people feel insecure). Both will depend on the technical implementation and are therefore impossible to test with a dummy.

Picture 5 *A prototyped interface, using a touch screen simulating an interface with hard and soft keys.*

•   Engineering prototypes: functioning prototypes of the final product, with real hardware and software. In this project the quality of the interface was improved in several small steps, involving users every time, in an almost RITE [4] approach. The products are reacting now on what users are doing and new input is found. The scroll wheel in the shown project was a novelty in these kind of products and created many problems as people did not expect it and therefore have problems finding it and understanding its purpose.
•   Beta test: testing within a chosen client environment. A test situation in a lab, with defined tasks, no stressful conditions, and an unusual setting is quite different from a large group of users performing their daily work in a real working situation, including stressful conditions [9]. This results mostly in several unexpected observations. In the beta test it became clear that the heavy usage of the product by large user groups (exceeding 80 different persons) resulted in a long list of users and jobs in the interface. Though the scroll wheel was designed to facilitate easy scrolling,

this exceeded our expectations by far, and the users started complaining about it. Even though this was not an item that came up during the usability testing before. On the other hand a feature Océ Touchto Print, was unanimous liked much by the users when introduced in the test sites. Fingerprint recognition is used to start printing the documents of a user straight away, no scrolling required.

An important advantage of this recurring way of user testing within a project is that user testing can really support interaction designers choosing e.g. conceptual directions. It is often not necessary to use very advanced prototypes to gain sufficient insight, as long as one remains aware of the limitations of a chosen test.

Picture 6. O*cé Touchto Print: usage of finger print recognition not only to identify a user, but the product directly prints the document the user did send at 72 pages per minute.*

6. Who will perform the evaluation

In deciding who will perform a usability test the objectivity and distance a person has about the product and (interaction) design is important. In the past we experienced the hard way what happens if a design team under stressful conditions has to perform usability test on their own concept: tunnel vision. Problems are simply not seen / heard or taken too lightly. The product went to market before we realised there was a mismatch between concept and some of our more traditional operators. We are all human after all.

Therefore Océ has chosen to separate the two roles: designer and usability engineer. A usability engineer is someone with specific knowledge about cognitive ergonomics, psychology, and methodology. The usability engineer does not design (within this role), but advises designers about what users want,

need or are capable of and validates concepts. Notice that these two roles can be fulfilled by one person, but never for the same product.

The test role can be fulfilled by both external parties or by in-house specialists. In literature can be seen that this issue divides the world of specialists in two, either explicitly in favour of the 'objective external' approach or the 'bring it in the team' approach [10].

At Océ we have experiences with both, but favour the last. Although it is beyond doubt that the person fulfilling this role needs to be objective, there is another aspect involved: convincing the design team that changes are necessary. And preferably without needing thick reports with loads of statistical calculations, as that takes far too much time. Best way to assure test results are leading to changes is to become a project member, for as long as the period required organising, doing and evaluating a usability test. In that period every one in the project will become involved, raise questions, helps to get the prototypes ready and –most important – become observers during the tests. In our tests project leaders, managers and product managers were observers next to the design team, to see themselves what the usability of their product is. With the result that appropriate steps are taken already to solve the reported problems, before the test report is finalised. Still, it is important that to remain objective it is a condition that whatever the test results are, the implications are not to be solved by the usability engineer. Better to remain the messenger.

Another reason why we favour in house usability engineers is that by doing so explicitly meta-knowledge across projects is gathered. Tests that are conducted are better and faster organised, the results are better understood thanks to a large number of similar tests before.

The experiences we had with external parties evaluating our product are a bit diverse. The great advantage is that these people are objective, the results of the evaluation have no effect on their work what so ever. In addition, an external party does not have (too much) inside knowledge. This can help us as a reality check, a mirror. Although we consider our products to be very important, for most of our users it is not the centre of their (working) lives. Sometimes you need to be reminded of that.

On the other hand the lack of contextual knowledge of our customers, users and their tasks is at the same time what makes it so difficult for an external party to organise a good usability test. Before they can perform the actual test, much time is required

to ensure they will conduct the right test, with right users. Another problem is convincing a project to take follow up steps in the development. After a test is performed and results are communicated, it is easy for a project to dismiss results with all kinds of arguments. It helps if the usability engineer is still around to convince a project to take follow up steps in the development.

Still we think an external party can actually be used successfully, when he adapts himself to the role of what an in-house usabilty engineer has, working within the team.

7. Conclusions

There is a gap between what good usability testing requires (time, sufficient real users, and stable prototypes) and the conditions imposed by the nature of product development, which is fuzzy and has an aspect of secrecy. As we define usability testing in the context of its contribution towards the final product, the challenge is to adapt usability testing to the nature of product development.

The main conclusions in this case study are:
(1) Usability testing must be performed within the time span of a project, when a design is not final yet, when there is ample time left after conducting the test to change things. If this implies that not all requirements can be met for a fully objective usability test, so be it.
(2) Usability testing must be done 'within' the project, with a project supporting test and test goals, with evaluators that are project people.
(3) To balance objective testing and convincing a project the usabilty engineer should be part of the project / design team for as long as the usabilty test requires. To ensure objectivity the roles of (interaction) designer and tester must be separated and the tester should not be the one who has to solve the implications of a test.
(4) Test only with real users if it is possible. Only when it concerns anthropometrical tests it is possible to use in house persons. Mostly 5- 6 test persons are sufficient.
(5) Beta tests on client sites are vital as a follow up of lab situation tests. The lab situation, with defined tasks and an unusual setting can result into different observations then in case of a reality check with real users, in a real working situation.

8. Discussion

The most important issue we haven't solved yet is what to do when it is impossible to use real users. We learned the hard way that you should be very careful to interpreted test results with Océ employees. How to

deal with this: focus groups? A way is to integrate a number of real users in the team some way or other [likewise 11] to be able to do test without the problem of secrecy. However, you are more or less training these customers as we perform tests often.

Last issue we are experimenting with is solving the lack of behaviour in the decisive dummy-tests. An option is to use interface prototypes on an existing product that really functions. But looks, feels and behaviour are different from the final product, so how to interpreted this?

References
[1] http://www.oce.com/
[2] www.iea.com
[3] Dumas, J.S. and Redish, J.C. *A Practical Guide to Usability Testing revised edition* (1999), Intellect books, ISBN: 1841500208
[4] Medlock M.C., Wixon D., Terrano M., Romero R.L., Fulton B. *Using the RITE method to improve products; a definition and a case study,* Usability Professionals Association (UPA2002), Orlando, July 2002.
[5] Nielsen, J., and Landauer, T. K. 1993. A mathematical model of the finding of usability problems. Proceedings ACM/IFIP INTERCHI'93 Conference (Amsterdam, The Netherlands, April 24-29), 206-213
[6] Woolrych, A. and Cockton, G., *Why and When Five Test Users Aren't Enough*, in Proceedings of IHM-HCI 2001 Conference: Volume 2, eds. J.
[7] Spool, J. and Schroeder, W. *Testing Websites : Five Users is Nowhere Near Enough.* In Proc. *CHI 2001, Extended Abstracts,* ACM 285-286
[8] Snyder C., *Paper prototyping: the fast and easy way to design and refine user interfaces* (2003), ISBN 1-55860-870-2
[9] Merholz, P.*Get Out of Your Lab, and Into Their Lives* (2006) http://www.adaptivepath.com/publications/essays/archives/000569.php
[10] Becker L. *90% of All Usability Testing is Useless*, June 16, 2004, www.adaptivepath.com/publications/essays/archives/000328.php
[11] Mulhern T and Lanthrop D. *Building and tending bridges: rethinking how consultants support change.* DMJ summer 2003, 27-33