

## Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications

Raptis, Dimitrios; Papachristos, Eleftherios; Kjeldskov, Jesper; Skov, Mikael B.; Avouris, Nikolaos

*Published in:*  
Behaviour and Information Technology

*DOI (link to publication from Publisher):*  
[10.1080/0144929X.2013.848239](https://doi.org/10.1080/0144929X.2013.848239)

*Publication date:*  
2014

*Document Version*  
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Raptis, D., Papachristos, E., Kjeldskov, J., Skov, M. B., & Avouris, N. (2014). Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications. *Behaviour and Information Technology*, 1168-1179. <https://doi.org/10.1080/0144929X.2013.848239>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.



# Studying the effect of perceived hedonic mobile device quality on user experience evaluations of mobile applications

When people interact with digital artifacts they perceive their *pragmatic* and *hedonic qualities*. In the case of interacting with mobile devices and applications users seek utility as they try to satisfy certain needs, but at the same time they have certain feelings and emotions when, for example, they feel attached to their personal phone and/or trust its brand. Due to this strong relation between users and mobile devices a significant problem occurs when researchers want to evaluate the user experience of a mobile application in laboratory settings: the selection of an appropriate mobile device. Towards this end, this paper aims to unveil the effect of perceived *hedonic quality* of a mobile device on the user experience evaluation results of an application. Our results show that the perceived *hedonic quality* of a mobile device significantly affected the perceived *pragmatic* quality of the application, but not the *hedonic* one.

Keywords: mobile devices, mobile applications, user experience, laboratory settings, hedonic quality, pragmatic quality, attractiveness, brand.

## 1 Introduction

The diffusion of mobile technologies has been considerable over the past years. Today, people constantly use their personal mobile devices and take advantage of the hundreds existing applications to accomplish everyday tasks, communicate, play, socialize, etc. During the last decade, the mobile phone has rapidly transformed from a single purpose device, mainly used for making phone calls, to a powerful digital artifact. In addition, since interaction with mobile devices becomes more widespread, researchers' perception regarding the user is evolving. During the 70's the user was understood as a gear in a rational machine, during the 80's as a source of error, during the 90's as a social actor, and now as a consumer (Kuuti, 2001). When such users/consumers want to buy a mobile device they face a vast selection of candidate devices that offer more, or less similar functionality and have differences in their style, design, brand, operating system, etc. Acting as consumers many users don't treat the device they have

intentionally selected just as a tool that helps them to deal with various tasks, thus focus only on utility, but as something more (Coates, 2002). For them the mobile device is a *product* that was intentionally chosen among others, it has significant *value* and *meaning* (Hallnäs and Redström, 2002) and it projects elements of their personality (Jordan, 1997; Phau and Lau, 2001). We may also find consumers that move a step further and treat their devices as something so important that needs to be personalized and made more attractive according to their personal taste and style. Personalization can occur at the software level, affecting how usable a product is perceived (Tossell et al., 2012), or at the hardware level and therefore we may find: users that customize their devices in various ways (Moggridge, 2007; Marathe and Sundar, 2011), special editions of a particular device, or even large fashion industries involved in the production of a device (for example Motorola Razr V3i D&G handset).

At the same time, researchers and practitioners that are involved in developing new mobile applications face the challenge of properly evaluating their prototypes by measuring the users' experience with them (Hassenzal and Tractinsky, 2006). There is a variety of user experience evaluation methods available (Vermeeren et al., 2010), but especially in the case of evaluating mobile applications in laboratory settings researchers and practitioners need to decide about the:

- Reasons behind conducting an evaluation. Thus, they need to decide on how to link the evaluation and the design process and how to use the findings in order to improve their prototype (Hornbæk, 2010).
- Participants: how many participants are going to be involved, what kind of previous experience with the application they should have, what kind of general experiences might affect their perception for the application (e.g. internet usage).

- Evaluation instruments: what needs to be measured (usability, pleasure, user satisfaction, etc.) and how it is going to be measured (direct observations, task completion times and rates, questionnaires, etc.).
- Tasks: select between *goal mode* and *action mode* (Hassenzahl, 2003; van Schaik et al., 2012). In goal mode participants are provided with specific tasks and they try to be as effective and efficient as possible, since the fulfillment of the goal becomes the core of their interaction. On the contrary, in action mode the action itself is to the fore and using the mobile application is at the center of their interaction.
- Mobile devices: which mobile devices are going to facilitate the mobile application during the laboratory evaluations.

If we focus on the decision for the selection of an appropriate evaluation instrument, then we observe that user experience can be measured through a variety of questionnaire constructs such as hedonic and pragmatic qualities, fun, flow, enchantment, etc. Despite the fact that there is a significant amount of research directed towards understanding these constructs, there still are challenges that HCI needs to overcome. The most important one is derived from the fact that we do not know if these constructs are enough to describe user experience because the relations among them have not been studied in detail (Bargas-Avila and Hornbæk, 2011). Thus, the challenge for researchers and practitioners, when they want to make a laboratory evaluation with questionnaires, is to decide which of the user experience constructs they will use in order to evaluate their applications. Furthermore, in relation to the selection of a mobile device for the evaluation, an important challenge is that practitioners and researchers do not know the “complete” digital artefact of their users. If for example they are developing an Android<sup>TM</sup> application, then they know the software part of the artefact,

but they have to deal with a huge variety at the hardware part. The reason for this variety is that there are many Android<sup>TM</sup> manufacturers that produce devices with different characteristics, for example different screen sizes, weight, colours, materials, etc. These characteristics, which for some researchers may seem of secondary importance in the context of mobile user experience evaluations, are treated as very significant from potential mobile phone consumers (Ling et al., 2007) and can determine the commercial success, or not of a specific device (Kim et al., 2012).

## 2 Related Work

In this paper we will focus on the cases where the user experience of a mobile application is measured through questionnaires in laboratory settings and the participants first interact with the application and then they are asked to formulate a judgment. There are two different approaches on how this judgment is shaped (van Schaik et al., 2012). The first approach is through *induction* (Kardes et al., 2004) and it implies that the participants are building an overall assessment by carefully considering, analyzing and weighting all the relevant attributes (for example usability, aesthetics, functionality, etc.). The second approach, which is more recent one and supported by a large amount of empirical evidence, proposes that participants are using simple cognitive rules and *infer* a judgment based on the availability of information (Kruglanski and Cigerenzer, 2011). An example from the HCI domain is that participants infer that a product will be usable because it is appealing.

On the other hand, when users interact with a digital artifact their judgment is affected by *integral* and *incidental affect* (Hazzenzahl and Monk, 2010; Hassenzahl et al., 2010; van Schaik et al., 2012). *Integral affect* are the feelings produced and attributed to the artifact and *incidental affect* are the feelings produced by

circumstances. These feelings contribute to the formulation of judgments regarding the perceived *hedonic* and *pragmatic quality* of the artifact (Hassenzahl, 2004). *Pragmatic quality* (i.e. perceived usability) refers to a judgment in relation to the possibility the artifact to support participant's "do-goals" and *hedonic quality* is a judgment to the artifact's ability to facilitate pleasure while using it (Hassenzahl et al., 2010). These judgments are also affected by the availability of information regarding the artifact's features. If for example users own an artifact then they know most of the artifact's features and thus *integral affect* is highly influencing their judgments. On the other hand, in the context of evaluating a mobile application in laboratory settings the participants are not familiar with all the application's features as they do not have enough time to fully analyze it, and as a result they *infer* judgments for the *hedonic* and *pragmatic* quality of the application based either on previous experience, or specific artifact features that are easy to perceive (such as the attractiveness of the application).

In the context of mobile evaluations the effect of specific features to user experience have been studied in detail. For example, in relation to perceived *pragmatic quality* of a mobile device we may find studies that compare the performance of various input methods (for example, types of keyboards: Clarkson et al., 2005). On the other hand regarding *hedonic quality* we may find evidence that attractive things work better (Norman, 2004) and studies that specifically investigate the notion of attractiveness in relation to perceived usability (Tractinsky et al., 2000; Chawda et al., 2005). In line to these findings are also recent studies that unveiled the effect of mobile phone's attractiveness on effectiveness and efficiency (Quinn and Tran, 2010; Sauer and Sonderegger, 2010). Attractiveness has also been studied in relation to the first impression and the impression after usability testing, both in the context of websites (Lindgaard et al., 2006; Lindgaard et al., 2011) and mobile computing (Sauer and

Sonderegger, 2010), and it showed that the judgment regarding the attractiveness of a digital artifact is mostly shaped by the first impression. Additionally, Ling et al. (2007) examined the impact of specific mobile phone characteristics to user satisfaction, showing that users' general attitude towards a mobile device can be explained by their preference towards specific characteristics, such as physical appearance and body color. Finally, there are studies that focus on the effect of brand during usability testing sessions (Bolchini et al., 2009; De Angeli et al., 2009).

### 3 Aim of this study

In this paper we focus on the evaluations of mobile applications in laboratory settings. In these cases the users' experience is influenced both by the mobile device and the mobile application. Furthermore, since in such evaluations participants are asked to focus on the application and formulate a judgment about it, the mobile device they use is treated as the means to conduct the task(s) they have to accomplish. As a result, in laboratory evaluations the perceived *hedonic* and *pragmatic* quality of the device contributes to participant's *incidental affect* and the perceived *hedonic* and *pragmatic* quality of the application contributes to participant's *integral affect*.

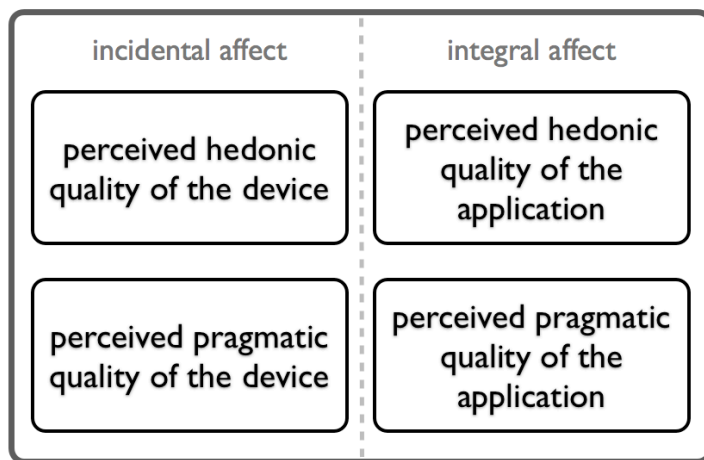


Figure 1. Perceived hedonic and pragmatic quality of a mobile device and a mobile application in the context of user experience laboratory evaluations.

In order to see how the HCI community deals with the *incidental affect* created by a mobile device in laboratory evaluations, we reviewed relevant papers in two CHI conferences (2010 and 2011). Relevant were the papers that included the words ‘mobile device’ or ‘smartphone’ or ‘PDA’ and a mobile application was involved. In about 100 papers researchers: a) simply mention the model of the used mobile device without arguing why they chose this device (for example Bidwell et al., 2010), b) state that they used a mobile device of a certain operating system without specifying a model (for example Durrant et al., 2011), c) use various mobile devices and sum up the results (for example Quinn and Tran, 2010), d) state that they selected a mobile device based on specific needs on hardware and/or software (for example Costanza et al., 2010), or e) do not clearly present any details about the device (for example McLoughlin and Ciolfi, 2011). We identified just one case where the researchers tried to minimize the effect of mobile device hedonic features by using custom made mobile device casings to avoid influencing their participants (Chong and Gellersen, 2011).

Consequently, in the vast majority of these papers researchers do not present a clear argumentation on why they selected a specific device and they also tend to ignore the possible effect of the perceived *hedonic* and *pragmatic* device quality (incidental affect) on the evaluation results of their mobile applications. We will illustrate why we believe this approach can be misleading through an analogy: Treating the perceived *hedonic* quality of a mobile device as a parameter that has minor effect on user experience is like assuming that drivers of different cars will experience the same road in a similar way, if their cars facilitate the same perceived *pragmatic* quality (same engines, same horsepower etc.), independently from their cars’ perceived *hedonic* quality (for example brand, visual design, etc.).

Therefore, in this paper we are interested in this research question: “*Does the incidental affect created by a mobile device’s perceived hedonic quality has an effect on the user experience evaluation results of a mobile application in laboratory settings?*”

The rest of the paper reports from an experiment we conducted in order to answer this research question. Through this experiment we discovered that the perceived *hedonic quality* of a mobile device has a significant effect on users’ experience with a mobile application. The rest of the paper is organized as follows. First, we present our method and the experimental conditions. Then, we outline our findings and discuss these findings against related research. Finally, we conclude on our work by presenting our suggestions for the reasons that have created these results.

## **4 Method**

In order to answer the research question we initially made two pilot studies to test our experimental setup and then we adopted a between-group experimental design (Lazar et al., 2010) and asked two groups of users (each group interacted with one device) to interact with and then evaluate the same application on two different devices. In the following subsections we present our key decisions.

### ***4.1 Device selection***

One key decision in our experiment was the selection of devices. In our selection process we had to make sure both that the two devices would offer as similar perceived *pragmatic* qualities as possible and that the perceived *hedonic* qualities would have clear and distinct differences. We argue that our experimental design shares common ground with studies where *framing* on user judgments is under investigation (for example Hartmann et al., 2008). In such studies positively and negatively framed statements/questions of the same *meaning* are presented to participants and then they

are asked to make a judgment. In our case we wanted to provide the same “*meaning*” (same application with the same functionality and utility on both devices) framed by different perceived *hedonic qualities* of the devices. In order to achieve a big distance between the perceived *hedonic qualities* we had to select two devices with clear and notable differences on style, visual design, materials used for the casing and age. We chose these device features as the base to frame our study, by considering previous studies that identified which mobile features users consider as important (Ling et al. 2007). Having these features as a starting point we chose to include to the experiment an iPod Touch<sup>TM</sup>, 2<sup>nd</sup> generation, A1288 model (Device A), which when the study took place it was recently introduced to the Greek market and a Dell X51v<sup>TM</sup> (Device B) that even then it was considered an old PDA with somehow outmoded design (Figure 2). This specific device selection allowed us to have a big distance on the perceived *hedonic quality* and thus a clear *framing* for the *incidental affect* created by their *hedonic qualities*.

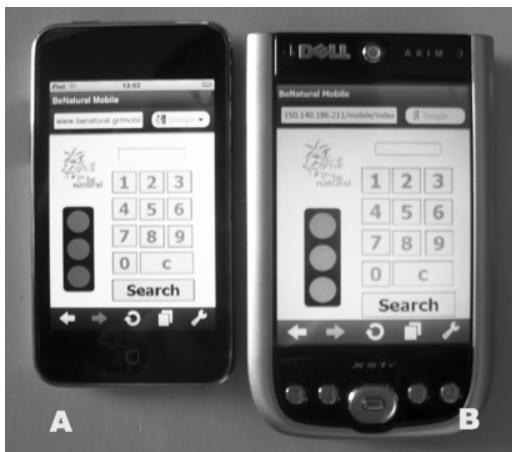


Figure 2. Test application on Devices A and B.

#### ***4.2 Application selection***

Any mobile application would have been suitable for our experiment since our purpose was not to evaluate the application per se, but to investigate the effect of perceived

*hedonic* mobile device quality on its evaluation results. We chose to use an application called *beNatural* as a test application mainly because it was developed by us and therefore: a) none of the participants had any prior experience with it, and b) we could easily make the necessary software changes. By experimentally controlling for prior experience with the test application we have also eliminated its effect on perceived usability as identified by Quinn and Tran (2010) and McLellan et al. (2012).

In short, *beNatural* (Figure 2) is an application that allows customers to be informed about the environmental impact of a product prior to purchasing it (for example to know if a product is made from recyclable materials). Information about products is contributed to the system's database by the users through a desktop website, following the Wikipedia example. Inquiries about a product are conducted through the mobile application by entering the product's barcode through a numerical keypad (Figure 2). The system's response about a particular product is presented to the users through a traffic light metaphor (Figure 2). Red light indicates that a product should not be bought, yellow light that it could be avoided, green that is suitable to buy it and when all lights are off then there is no information in the database. The responses are personalized according to each user's profile by applying a filtering mechanism, which describing it is out of the scope of this paper.

#### ***4.3 Experimental setup***

A series of actions were taken to ensure that both groups were offered similar perceived *pragmatic* qualities for the device and the test application and similar *hedonic quality* for the application. Since we wanted our software to look and feel the same in both devices we installed the Opera Mini<sup>TM</sup> browser (Figure 2) and we modified the interface of the test application in order to achieve that the size of the application and the size of

the on-screen buttons would be as similar as possible in both cases, despite the devices' differences on screen size and resolution. Additionally, we decided that our subjects should interact with the test application in a unified way and thus the group that interacted with Device B was not provided with a stylus and we locked Device A resize capabilities. Consequently, all participants experienced the test application by using their fingers and interacted with a similar interface.

A pilot study was conducted to check if the participants could perceive the big differences in devices' hardware (processor, screen responsiveness, Wi-Fi speed, etc.). Four participants were asked to insert a barcode as fast as they could 15 times, repeatedly on both devices (in total 60 repetitions on each device). For each effort the timer was started when they pressed the first barcode digit and it was stopped when the system responded for a product. A Latin Square design (Edwards, 1951) was adopted in order to avoid first-order and carry-over effects and the barcode they were asked to enter was "1234567890" in order to minimize cognitive load.

All four participants took similar time to accomplish the task in both devices. The average time to insert one barcode and see a response and the average error rate in 60 repetitions were: 10.39sec (SD=0.748) and 1.5errors (SD=0.911) for Device A, and 10.31sec (SD=0.581) and 1.63 errors (SD=0.758) for Device B. The average times were checked for differences between the two devices using T-Tests and there were no statistically significant differences. These minimum differences on insertion-response time and error rate showed that our participants could achieve the same goal at almost the same time, independently from the device they used and that the differences in hardware were not perceived, mainly because the test application was not resource demanding and the task was relatively simple.

In order to have accurate data on the perceived *hedonic quality* differences between the two devices, we asked 15 additional participants to use an online form that contained one image for each device (the devices were depicted from the front and their screens were turned off) and formulate a judgment for the devices' *hedonic* and *pragmatic quality* (AttrakDiff2: Hazzenzahl and Monk, 2010; van Schaik et al., 2012) and *attractiveness* (Quinn and Tran, 2010). The sequence the devices were presented was randomized.

Results showed that *hedonic quality* was perceived on average  $M=5.16$  ( $SD=0.96$ ) for Device A and  $M=3.31$  ( $SD=1.45$ ) for Device B, and this difference was significant ( $t(28)=4.105$ ,  $p<.001^{**}$ ) according to a two-sample T-test. *Attractiveness* was rated  $M=3.66$  ( $SD=0.71$ ) for Device A and  $M=2.43$  ( $SD=0.84$ ) and this difference was also significant ( $t(28)=-4.303$ ,  $p<.001^{**}$ ). The difference in the *pragmatic quality* was not significant ( $t(28)=2.041$ ,  $p=.052$ ) and it was perceived on average  $M=4.81$  ( $SD=1.39$ ) for Device A and  $M=3.98$  ( $SD=0.73$ ) for Device B.

Consequently, the results from the two pilot studies showed that our experimental setup offered both groups similar *pragmatic quality* and different *hedonic quality* for the device, and similar *pragmatic* and *hedonic qualities* for the application.

#### **4.4 Measures**

Two different types of measures were used during the two laboratory evaluation sessions that we conducted next: a demographics questionnaire and an evaluation questionnaire. We created our own version of the demographics questionnaire, but we chose to use existing, established questionnaires for evaluating the participants' experience with the application. Details about the questionnaires are presented next.

#### 4.4.1 Demographics questionnaire

The demographics questionnaire comprised 10 questions that documented the participants' experience with Devices A and B and general aspects like internet usage, familiarity with mobile devices, etc.

#### 4.4.2 Evaluation questionnaire

The used evaluation questionnaire was created by combining three established questionnaires: USE (Lund, 2001; Tullis and Albert, 2008), AttrakDiff (Hassenzahl, 2004) and "Pleasure while interacting with products" (Jordan, 2000) questionnaires. The produced evaluation instrument contained 58 items in total and it allowed us to obtain a holistic view on how our participants experienced their interaction with the test application.

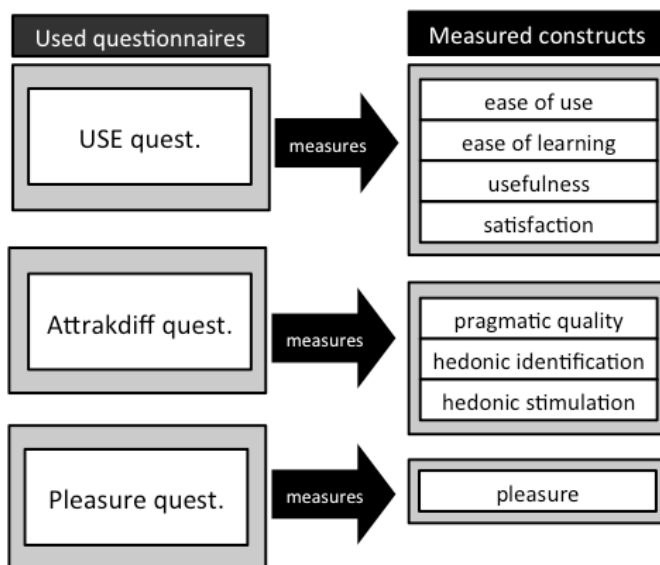


Figure 3. Measured evaluation constructs.

Figure 3 depicts the evaluation constructs measured by the evaluation questionnaire. We point out that some of these constructs are either highly correlated, or even measuring similar aspects (for example *easy of use* and *ease of learning* are highly

correlated to *pragmatic quality*). Nevertheless, we chose to include all these constructs in order to have multiple data sources.

The following subsections present in detail the user experience constructs that were measured by the evaluation questionnaire.

#### 4.4.2.1 Perceived Usability

We measured perceived usability with USE questionnaire (Lund, 2001; Tullis and Albert, 2008). USE measures perceived usability through *usefulness* (7 items), *ease of use* (4 items), *ease of learning* (3 items) and *satisfaction* (7 items). Participants were provided with statements and they had to formulate their judgment on a 7-point scale (strongly agree-strongly disagree).

#### 4.4.2.2 Perceived Hedonic and Pragmatic quality

*Hedonic* and *pragmatic qualities* were measured by AttrakDiff (Hassenzahl, 2004). AttrakDiff (Hassenzahl, 2004) is a questionnaire that presents pairs of opposite adjectives, used in this case on a 7-point scale. *Pragmatic quality* (7 items) is related to perceived usability and it indicates if users can fulfill their goals while using a product. *Hedonic quality* is further decomposed to *hedonic identification* (7 items) and *hedonic stimulation* (7 items). *Hedonic identification* measures if users identify with a product and if it can communicate their personal values. *Hedonic stimulation* indicates how challenging and novel a product is perceived by the users. AttrakDiff contains also two items related to the user's overall impression towards an interactive product: *bad/good* and *ugly/beautiful*.

#### 4.4.2.3 Pleasure

Finally, Jordan's *pleasure* questionnaire (Jordan, 2000) indicates to what extent users

would characterise their interaction with a product as pleasurable. *Pleasure* was measured by 14 items, on a 7-point scale (strongly agree-strongly disagree).

## **5 Experiment**

### ***5.1 Participants***

54 additional individuals, 30 men and 24 women, aged 22 to 35 (M=25.1) participated in the experiment. From them, 3 were excluded from data analysis as they partially completed the evaluation questionnaire. All of them were undergraduate and postgraduate students from four University of Patras departments in Greece (Departments of Computer Science, Chemistry, Biology, and Educational Sciences and Early Childhood Education). All of them volunteered for the experiment.

### ***5.2 Procedure***

Two weeks prior to the experiment all participants were asked to fill in the online demographics questionnaire. From the demographics questionnaire we found out that none of the subjects had any experience with Device B (Dell X51v<sup>TM</sup>) and six of them had limited experience with Device A (iPod Touch<sup>TM</sup>) and/or similar products (for example iPhone<sup>TM</sup>). These six participants were equally distributed to the two groups and all the rest were assigned randomly. 27 participants interacted with Device A and 24 interacted with Device B.

The experiment evolved in three phases, all conducted in the same room in a usability laboratory and each subject participated individually. The initial phase was a five-minute general introduction to the test application, conducted by the same researcher who made sure that all participants heard the same narrative. In short, participants were informed about the basic services offered by the test application, the

type of tasks that can be performed on its mobile and the desktop parts and how users can benefit from the test application in general. Since we wanted to focus on the application and not on achieving specific goals we decided to opt for the *action mode* (Hassenzahl, 2003; van Schaik et al., 2012) and we did not provide any tasks to the participants. Therefore, after the introduction each participant interacted freely with the test application for fifteen minutes.

In both groups the application was running in full screen mode to prevent the participants from interacting with the operating system. In order to mimic real world conditions seven products that are common in a supermarket were placed on a shelf next to them. Prior to each session the researcher made sure that five products were always present in the test application's database and two did not exist in order to give participants the possibility to insert a product to the system if they decided to do so. The selection of the two non-existing products was made randomly. All participants inquired for information for at least four products. Help was provided only when asked and only two participants asked for it. The average error rate in inserting barcodes was: 1.16 errors (SD=0.90) for Device A and 1.28 errors (SD=0.94) for Device B, without any significant effects. In comparison to the pilot study the participants made less errors (probably due to the fact that they did not have to be fast) and their average error rate was similar in both groups. None of the participants used the mobile devices' hardware buttons, nor interacted with the operating system.

At the final phase of the experiment participants were instructed to formulate a judgment about their experience with the test application by filling in the evaluation questionnaire.

## 6 Data analysis

As a first step in the data analysis all participants' ratings on the three questionnaires were combined and used as variables in an exploratory factor analysis. The aim was to reconcile for overlapping attributes, considering that the three questionnaires that were used assessed similar constructs (e.g. pragmatic quality and perceived usability), and obtain a reduced number of factors that would hopefully reflect new perceived *hedonic* and *pragmatic* qualities for the application.

After conducting the exploratory factor analysis, the second step of data analysis was to analyse each questionnaire individually, and we tested both groups for statistically significant differences using two sample T-Tests ( $p < .05$ ). T-Tests were used since we had one independent variable with two groups (devices A and B) and many dependent ones (the 8 measured constructs). A possible alternative would have been to conduct MANOVA, but this would mean loss of statistical power since DV's are highly correlated to each other. Another approach would have been to adopt a more conservative alpha ( $p < .01$ ), but this could lead to type II error. Therefore we decided to perform T-Tests (Saville, 1990) and report the  $t$ ,  $p$  and  $\alpha$  values. Furthermore, for each evaluation construct we conducted a two-way analysis of variance (ANOVA) with device as between groups factor and age and gender as within subjects factor. No significant interaction effects were found. Apart from the actual evaluation questionnaire participants were provided with an area for comments and none of them reported anything related to the mobile device.

## 7 Results

### 7.1 Exploratory Factor analysis

Our dataset did not allow for an item level response factor analytic procedure as the accumulated number of single items (58) exceeded the number of total observations

(51). Therefore the analysis was conducted on the eight evaluation constructs (*ease of use, ease of learning, satisfaction, usefulness, hedonic identification, hedonic stimulation, pragmatic quality and pleasure*). The Monte Carlo simulations with data permutations that were conducted on this dataset indicated a two-factor structure as the most appropriate. Examination of factor loadings in the oblique rotated structure revealed that the two emerging factors resembled the pragmatic and hedonic dimensions. *Ease of use* (.951), *satisfaction* (.674), *ease of learning* (.662) and *usefulness* (.306) contributed on the first factor, while *hedonic identification* (.800), *hedonic stimulation* (.643) and *pleasure* (.609) contributed on the second one. *Pragmatic quality* contributed to both factors (.596 and .735 respectively).

Despite the fact that *pragmatic quality* contributed to both factors, due to its high correlation with perceived usability, we interpreted the two emerging factors as *perceived usability/pragmatic* and *hedonic quality* of the application and we named them as *new\_pragmatic* and *new\_hedonic* respectively. The factor scores of these qualities were used in a one-way ANOVA analysis with mobile device as a between subject factor. The results of this analysis showed a significant main effect of the mobile device on the *new\_pragmatic quality* ( $F(1,49) = 6.724, p = .013^*$ ). Contrary, a statistical significant main effect of the mobile device could not be shown for the *new\_hedonic quality* ( $F(1,49) = 1.947, p = .169$ ). On average, iPod Touch<sup>TM</sup> participants perceived the *new\_hedonic quality* of the test application quite high ( $M=0.17, SD= 1.02$ ), whereas X51v<sup>TM</sup> participants rated this lower ( $M=-0.19, SD=0.76$ ). The same was the case for the *new\_pragmatic quality* for iPod Touch<sup>TM</sup> participants ( $M=0.32, SD=0.88$ ) and X51v<sup>TM</sup> participants ( $M=-0.36, SD=0.99$ ).

## **7.2 Examining each questionnaire individually**

Figure 4 depicts the average scores from users' answers on four constructs of particular

interest and the results of the two sample T-tests (t, p and  $\alpha$  values). All the remaining measured constructs did not have any statistically significant differences. In the following subsections we present our findings for each questionnaire.

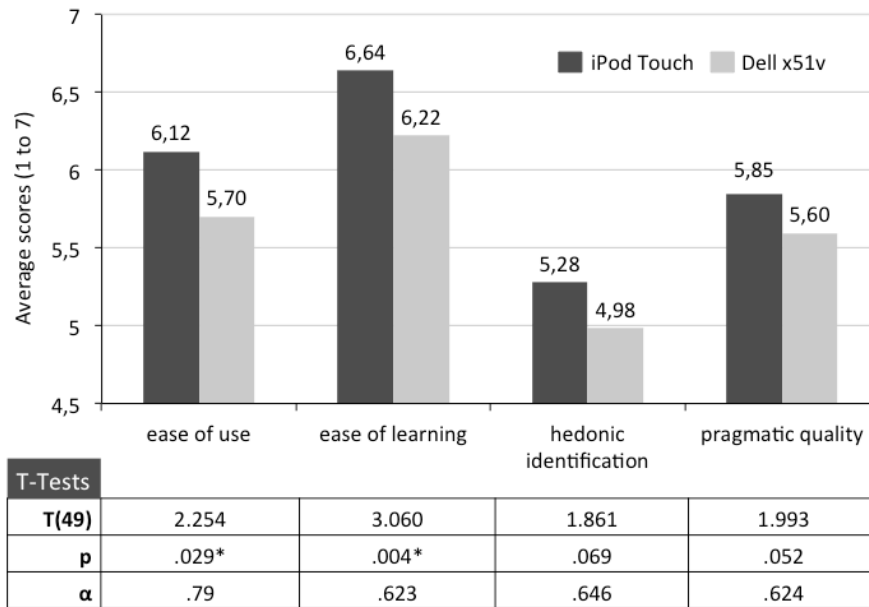


Figure 4. Average scores and results from two sample T-tests for four evaluation constructs with significant, or close to significant differences.

#### 7.2.1 Perceived usability of the mobile application (USE questionnaire)

The perceived usability of the application was measured through *ease of use* (Cronbach  $\alpha=.79$ ), *ease of learning* ( $\alpha=.623$ ), *usefulness* ( $\alpha=.611$ ) and *satisfaction* ( $\alpha=.731$ ) on a scale from 1 to 7. We found that the differences the two devices had on perceived *hedonic quality* affected significantly some of the usability related constructs of the application. On average, iPod Touch<sup>TM</sup> participants perceived *ease of use* of the test application quite high ( $M=6.12$ ,  $SD=0.63$ ), whereas X51v<sup>TM</sup> participants rated this lower ( $M=5.70$ ,  $SD=0.70$ ) and this difference, according to a two-sample T-test was significant ( $t(49)=2.254$ ,  $p=.029^*$ ). We further identified a difference for *ease of learning*, where iPod Touch<sup>TM</sup> users perceived the test application as very *easy to learn* ( $M=6.64$ ,  $SD=0.43$ ) and this was slightly better than X51v<sup>TM</sup> users ( $M=6.22$ ,  $SD=0.54$ ).

This difference was also significant ( $t(49)=3.060$ ,  $p=.004^*$ ).

### 7.2.2 Perceived hedonic and pragmatic qualities of the mobile application (AttrakDiff questionnaire)

Through Attrakdiff we measured *pragmatic quality* (Cronbach  $\alpha=.624$ ), *hedonic stimulation* ( $\alpha=.754$ ) and *hedonic identification* ( $\alpha=.646$ ). While our results on application's perceived *hedonic identification* and *pragmatic quality* showed higher values for iPod Touch<sup>TM</sup> participants, we identified no significant differences between the two groups of users even though we observed a trend both on *hedonic identification* ( $t(49)=1.861$ ,  $p=.069$ ) and *pragmatic quality* ( $t(49)=1.993$ ,  $p=.052$ ). iPod Touch<sup>TM</sup> users rated *hedonic identification* on average  $M=5.28$  ( $SD=0.62$ ) and X51v<sup>TM</sup> users  $M=4.98$  ( $SD=0.51$ ) and *pragmatic quality* was rated on average  $M=5.85$  ( $SD=0.48$ ) and  $M=5.60$  ( $SD=0.40$ ) respectively.

### 7.2.3 Experienced pleasure with the mobile application (Pleasure questionnaire)

We measured *pleasure* (Cronbach  $\alpha=.869$ ) using Jordan's questionnaire (Jordan, 2000) and we identified no significant differences between the two groups.

Overall, our results showed that the differences the two devices had on perceived *hedonic quality* affected significantly the perceived *pragmatic quality* of the application, despite the fact the devices offered similar perceived *pragmatic quality*. This result was observed when we examined the three evaluation questionnaires individually, as well as, when we conducted an exploratory factor analysis by combining the questionnaires.

## 8 Discussion

In relation to previous research work we believe that we moved a step further as we did not focus on studying the effect of the *hedonic* and *pragmatic qualities* of the application (as in Chawda et al., 2005), but we focused on the device. Additionally, we extended the study conducted by Quinn and Tran (2010) by focusing both on *hedonic* and *pragmatic* quality of the device and we eliminated the fact that they used many mobile devices and summed the results. Furthermore, we differentiate from the Sauer and Sonderegger (2010) study, as we chose to maximize the differences on the perceived *hedonic* device quality, by selecting two completely different devices and not two versions of the same phone. Finally, we differentiated from the related papers as we chose to include to our experiment a variety of evaluation questionnaires.

The most important finding of our study is that two user experience evaluations of the same application on two different devices provide significantly different results. In detail, data analysis showed that two groups of users interacting with the same application on devices that facilitate similar perceived *pragmatic* and different perceived *hedonic* qualities have different experiences. We were surprised to find out that the perceived *hedonic quality* of the application was not affected despite the fact that we *framed* the *incidental affect* by choosing two devices with big differences on perceived *hedonic* qualities. On the other hand, we observed a significant effect on the perceived *pragmatic quality* of the application.

We propose that the main reason that the devices influenced only the perceived *pragmatic quality* of the application and not the *hedonic* one is related to the type of the application we used in our experiment. Since the test application was mainly *pragmatic*, as its purpose is to assist users to decide about the environmental impact of a product, then the *incidental affect* that was created by the devices has affected only the user

experience constructs that the participants considered as relevant for the application. In other words the *hedonic quality* of the device influenced more the *pragmatic quality* of the application because the participants rated it as more important. We argue that a possible explanation to this situation is the “beauty dilemma”, which describes the phenomenon where potential users, even though they consider beauty as important, they discount beauty, visual design, and hedonic aspects, when it comes to the requirements of a new product, in favour of utility and usability (Diefenbach and Hassenzahl, 2009). Therefore we propose that our participants were incidentally influenced by the *hedonic quality* of the device and they transferred this effect only to the *pragmatic quality* of the application, due to “beauty dilemma”. Furthermore, the fact that we invited our participants for this experiment to a laboratory might urged them to act more rational (Diefenbach and Hassenzahl, 2009) and thus they unintentionally transferred the perceived *hedonic quality* of the device only to the constructs related to application’s perceived usability. Finally, another interesting fact about this finding is that we observed this phenomenon even though we opted for the *action mode* for the evaluation, and in these cases the *pragmatic* constructs do not have as significant impact as in *goal mode* (Hassenzahl and Ullrich, 2007). Therefore it will be interesting to find out if the same phenomenon will be observed when participants are asked to evaluate a leisure application and/or in conditions outside a laboratory. We believe that in this case the *hedonic quality* of the application will be influenced more since the participants will perceive it as more important. Of course, more research data are needed in order to have solid results.

In the following subsections we propose some possible explanations on which of the mobile device features influenced initially the perception regarding the *hedonic quality* of the device which then it influenced the perception of the *pragmatic quality* of

the application and we will discuss the implications of our findings on user experience evaluations of mobile applications in laboratory settings, in general.

### ***8.1 The effect of the brand***

One of the main factors that shape a consumer's opinion to choose a specific product among a set of possible alternatives is *brand* (Solomon et al., 2010). Furthermore, *brand trust* influences consumers to continue using a specific product, or to buy more items of the same brand (Delgado-Ballester and Munuera-Alemán, 2001). Considering the facts that mobile devices are products and that we chose to include in our experiment two of the most dominant and well-known brands in the area, it is safe to assume that brand might have played a role.

An experience with a brand can be characterized as *indirect*, created by marketing and advertisement, and *direct*, created by the actual interaction with a product (Rondeau, 2005). We believe that our participants had their own preconceptions about the brand of the device they used, even if, as documented by the demographics questionnaire, most of them had only *indirect* experiences with the devices (similar to an opinion that someone has about the brand of a very expensive car, which never drove), and these preconceptions affected the *pragmatic quality* of the application. As shown by Raita and Oulasvirta (2011) the effect of these preconceptions, or expectations towards a mobile device can be rather significant and strongly influence perceived usability. Thus, we believe that our users' attitude towards the brand might have imposed a halo effect (Nisbett and Wilson, 1997) that significantly influenced their judgment on the *pragmatic quality* of the application. Due to this halo effect users transferred positive traits associated with the brand of the Device A to the mobile application in a similar way that most humans consider beautiful as good (Dion et al., 1972). Thus, they transferred the fact that Apple<sup>TM</sup> is associated with easy to use and

learn products to their judgment of the mobile application (*ease of use* and *ease of learning* had statistically significant differences, Figure 4). Our argument is also further enhanced by a previous study, in the context of laptops, which affirmed that brand attachment affects the users' perception regarding perceived usefulness, beauty and pleasure (Tzou and Lu, 2009).

## **8.2 The effect of attractiveness**

Based on studies that show that *attractiveness* can possibly affect perceived usability (Tractinsky et al., 2000; Chawda et al., 2005; Quinn and Tran, 2010; Sauer and Sonderegger, 2010) we argue that the differences the two devices had in relation to attractiveness (pilot study results) influenced our participants, and for this reason the perceived *pragmatic quality* of the test application was significantly higher for the iPod Touch<sup>TM</sup> users. Thus the *attractiveness* of the device contributed to the perceived *hedonic quality* of the device and this perception was unintentionally transferred to the *pragmatic quality* of the application because it was considered as more important.

## **8.3 Implications to mobile user experience evaluations in laboratory settings**

The results of our experiment showed that users interacting with the same application on two devices with similar perceived *pragmatic* and different *hedonic* qualities had different experiences. Their experiences varied significantly on how they perceived the *pragmatic quality* of the application. Further studies are necessary in order to validate this finding and investigate if this result can be generalized and it is not circumstantial. Nevertheless, this experiment offers a strong indication that the *incidental affect* created by a mobile device during user experience laboratory evaluations, does affect the evaluation results of a mobile application. As a result, this finding may lead to a significant change on the way researchers set up and conduct their user experience

laboratory evaluations. Since the device can significantly affect the evaluation results of an application, the selection of an appropriate device becomes a decision as important as selecting the participants, deciding about the evaluation tools, etc.

A classification that categorizes mobile devices' characteristics has emerged from our effort to select two suitable devices for our experiment and it can act as a starting point for researchers and guide them on how to select an appropriate device for their user experience evaluations. We argue that each mobile device owns some *interaction*, *technical* and *physical* characteristics that shape the perceived *pragmatic quality* of the device, and some *intangible* and *physical* characteristics that influence the perceived *hedonic quality* of the device (Figure 5).

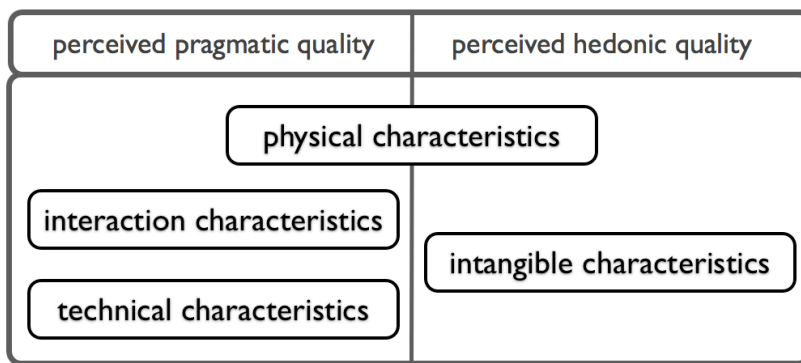


Figure 5. A proposed classification for the mobile device characteristics that influence the perceived hedonic and pragmatic quality of a mobile device.

*Technical* characteristics are related to the hardware components of the device and can have an effect on the quality of the experience. For example, when users are experiencing delays in a network connection, or in rendering mechanisms they can become frustrated. *Interaction* characteristics are related to the metaphors and the affordances (De Souza, 2005) facilitated by the device. Using a stylus or fingers, the size and the position of software and hardware buttons can be classified as *interaction* characteristics. *Physical* characteristics are the most difficult to categorize since they

can influence both *pragmatic* and *hedonic* quality. For example, the weight of a device can belong to both categories, but the used materials, the visual design, etc., are more related to *hedonic* quality. A detailed list of mobile devices' *physical* characteristics is presented by Han et al. (2004) who provided details on various hardware features and associated them with user satisfaction and by a more recent study that ranks these features in relation to how important users think they are (Ling et al., 2007). Finally, *intangible* characteristics are related to the added value and *intrinsic* meaning (McCarthy and Wright, 2004) of an experience. *Intrinsic* meaning refers to the value of an experience that is enjoyed for its own sake (i.e. appreciating a painting). Some of the *intangible* characteristics are brand, trust and attractiveness.

An example that can assist in better realizing this classification is the activity of driving a car. While driving users are perceiving the *pragmatic* quality of their car as they feel the power of the engine (*technical characteristic*), hold the leather steering wheel (*physical characteristic*), or operate the air-condition (*interaction characteristic*). At the same time, they are also perceiving its *hedonic* quality as they admire the beautiful design of the car's interior (*physical characteristic*), or feel younger and powerful inside their expensive convertible (*intangible characteristic*).

In our case this classification was helpful to provide the same perceived *pragmatic* and different *hedonic* quality for the mobile devices. We made sure that the differences in *technical* characteristics were not perceived, we offered similar *interaction* characteristics and ensured that the *pragmatic physical* characteristics were as similar as possible. Consequently, we believe that this classification can act as a starting point and guide researchers to select the appropriate device(s) when they evaluate the user experience of a mobile application in laboratory settings.

## 9 Conclusion

This paper shows that there are worth noting differences on users' judgments for the same application when they interact with devices that have different perceived *hedonic* and similar perceived *pragmatic* qualities, in the context of laboratory user experience evaluations. In our experiment the *incidental affect* that was shaped by the perceived *hedonic* quality of a mobile device, significantly affected the perceived *pragmatic* quality of the mobile application. Further studies are needed with different devices and applications in order to test if this phenomenon is circumstantial due to our experimental setup, or if there is a general *device effect* that significantly affects users' interaction with any mobile application.

Nevertheless, the results presented in this paper are worth reporting since they provide evidence that mobile devices play a more significant role than expected in shaping the evaluation results of a mobile application in laboratory settings. To use the car and road analogy, our results show that different cars do influence the opinion of a driver regarding the road he is driving, even if they facilitate the same *pragmatic* quality. The most important consequence of this finding is that yet another problem is added to every laboratory evaluation of a mobile application that is intended for a variety of devices: the selection of the device. We propose that researchers and practitioners should include to their mobile application evaluations as many candidate devices as possible since the impact of every device can not only significantly affect the users' judgments for a particular application, but is also more related, at least in this case, to the perceived *pragmatic* quality of the application. If it is not possible to include many devices then we propose that our classification of mobile devices characteristics can act as a starting point in order to select the appropriate one(s). Ideally, if researchers want to eliminate the effect of the device then they can, either select a neutral device in

terms of design and brand, or provide their subjects with devices with custom made design (as Chong and Gellersen, 2011) to influence less the evaluation results.

Furthermore, the most important finding of this study, besides the fact that the perceived device *hedonic* quality affected significantly applications' perceived *pragmatic quality*, is that it did not have an effect on its perceived *hedonic* quality. A possible explanation for this phenomenon is the fact that we used an application that its function was primarily perceived as *pragmatic*. We propose that it is really important for user experience research to investigate if the same phenomenon will appear when we use *hedonic* applications (for example leisure applications) since we believe that in these cases the perceived *hedonic quality* of an application will be affected more. Additionally, we need to further investigate the differences on users' judgments between *action* and *goal* mode (Hassenzahl and Ullrich, 2007) for these cases. We consider such research directions as important for HCI since we will gain a deeper understanding on the underlying mechanisms that our participants use in order to evaluate an application, we will understand how our experimental setups affect their judgments and this knowledge can contribute significantly on the user experience evaluation methods we apply.

As a future direction we intend to continue experimenting with different combinations of hardware and software in order to check the generalizability of our findings. Additionally, we plan to move one step further and check for possible overlaps and correlations between specific mobile device characteristics (such as device weight and screen size) and user experience constructs.

Acknowledgements: We would like to thank all the participants and Rondiri, F., Rontogiannis, N., and Stoica, A. for their effort to develop the test application. The research behind this paper was partly financed by the Danish Research Councils (grant number 09-065143).

## 10 References

- Bargas-Avila, J.A., and Hornbæk, K. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 2689-2698.
- Bidwell, N.J., Reitmaier, T., Marsden, G., and Hansen, S., 2010. Designing with Mobile Digital Storytelling in Rural Africa. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 1593-1602.
- Bolchini, D., Garzotto, F., and Sorce, F., 2009. Does Branding need web usability? A Value-Oriented Empirical Study. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09)*, 24-28 August Uppsala. Berlin: Springer-Verlag, 652-665.
- Chawda, B., Craft, B., Cairns, P., Rüger, S., and Heech, D., 2005. Do attractive things work better? An exploration of search tool visualizations. In *Proceedings of 19th BCS Conference on Human Computer Interaction: Volume 2 (British HCI '05)*, 5-9 September Edinburgh. BCS Press, 46-51.
- Chong, M.K. and Gellersen, H., 2011. How Users Associate Wireless Devices. *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 1909-1918.
- Clarkson, E., Clawson, J., Lyons, K., and Starner, T., 2005. An empirical study of typing rates on mini-QWERTY keyboards. In *CHI '05 extended abstracts on Human factors in computing systems (CHI EA '05)*, 2-7 April Portland. New York: ACM Press, 1288-1291.
- Coates, D., 2002. *Watches tell more than time: Product design, information and the quest for elegance*. New York: McGraw-Hill.
- Costanza, E., Panchard, J., Zufferey, G., Nembrini, J., Freudiger, J., Huang, J., and Hubaux, J.P., 2010. SensorTune: a Mobile Auditory Interface for DIY Wireless Sensor Networks. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 2317-2326.

- De Angeli, A., Hartmann, J., and Sutcliffe, A., 2009. The effect of brand on the evaluation of websites. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09)*, 24-28 August Uppsala. Berlin: Springer-Verlag, 638 – 651.
- De Souza, C.S., 2005. *The semiotic engineering of Human-Computer Interaction*. Cambridge, Massachusetts: MIT Press.
- Delgado-Ballester, E. and Munuera-Alemán, J.L., 2001. Brand trust in the context of consumer loyalty. *European Journal of Marketing*, 35 (12), 1238 – 1258.
- Diefenbach S. and Hassenzahl, M., 2009. The "Beauty Dilemma": Beauty is valued but discounted in product choice. In *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, 4-9 April Boston. New York: ACM Press, 1419-1426.
- Dion, K., Berscheid, E. and Walster, E., 1972. What is beautiful is good. *Journal of Personality and Social Psychology*, 24 (3), 285-290.
- Durrant, A., Rowland, D., Kirk, D.S., Benford, S., Fischer, J. and McAuley, D., 2011. Automics: Souvenir Generating Photoware for Theme Parks. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 1767-1776.
- Edwards, A.L., 1951. Balanced Latin-Square Designs in Psychological Research. *The American Journal of Psychology*, 64 (4), 598-603.
- Hallnäs L. and Redström J., 2002. From use to presence: On the expressions and aesthetics of everyday computational things. *Trans. on Computer-Human Interaction*, 9 (2), 106 – 124.
- Han, S.H., Kim, K.K., Yun, M.H., Hong, S.W. and Kim, J., 2004. Identifying mobile phone design features critical to user satisfaction. *Human Factors and Ergonomics*, 14 (1), 15-29.
- Hartmann, J., De Angeli, A. and Sutcliffe, A., 2008. Framing the User Experience: Information Biases on Website Quality Judgement. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*, 5-10 April Florence. New York: ACM Press, 855-864.
- Hassenzahl, M., 2003. The Thing and I: Understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A.F. Monk, and P.C. Wright (eds), *Funology: From Usability to Enjoyment*. Dordrecht: Kluwer Academic Press, 31-42.

- Hassenzahl, M., 2004. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19 (4), 319–349.
- Hassenzahl, M., and Tractinsky, N., 2006. User Experience – a Research Agenda. *Behaviour & Information Technology*, 25 (2), 91-97.
- Hassenzahl, M., and Ullrich, D., 2007. To do or not to do. Difference in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19 (4), 429-437.
- Hassenzahl, M., and Monk, A., 2010. The inference of Perceived Usability from Beauty. *Human-Computer Interaction*, 25 (3), 235-260.
- Hassenzahl, M., Diefenbach, S., and Göritz, A., 2010. Needs, Affect and Interactive Products – Facets of User Experience. *Interacting with Computers*, 22 (5), 353-362.
- Hornbæk, K., 2010. Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology*, 29 (1), 97-111.
- Jordan, P.W., 1997. Products as personalities. In M.A. Hanson, ed. *Contemporary Ergonomics 1997*. London: Taylor & Francis, 449-453.
- Jordan, P.W., 2000. *Designing Pleasurable Products. An Introduction to the New Human Factors*. London: Taylor & Francis.
- Kardes, F.R., Posavac, S.S., and Gronley, M.L., 2004. Consumer inference: A review of processes, bases and judgment inference. *Journal of Consumer Psychology*, 14, 230-256.
- Kim, K., Proctor, R.W. and Salvendy, G., 2012. The relation between usability and product success in cell phones. *Behaviour & Information Technology*, 31 (10), 969-982.
- Kruglanski, A.W., and Gigerenzer, G., 2011. Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118 (1), 97-109.
- Kuutti, K., 2001. Hunting for the lost user: From sources of errors to active actors and beyond. In *Proceedings of Cultural Usability Seminar*, 24 April Helsinki. Available from: [http://mlab.uiah.fi/culturalusability/papers/Kuutti\\_paper.html](http://mlab.uiah.fi/culturalusability/papers/Kuutti_paper.html) [Assessed 19 July 2012]
- Lazar, J., Feng, J.H., and Hochheiser, H., 2010. *Research methods in human-computer interaction*. West Sussex: Wiley & Sons.

- Lindgaard, G., Fernandes, G., Dudek, C. and Brown, J. Attention web designers: You have 50 milliseconds to make a good first impression!. *Behaviour & Information Technology*, 25 (2), 115-126.
- Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P., 2011. An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. *ACM Transactions on Computer-Human Interaction*, 18 (1), Article No. 1.
- Ling, C., Hwang, W. and Salvendy, G., 2007. A survey of what customers want in cell phone design. *Behaviour & Information Technology*, 26 (2), 149-163.
- Lund, A., 2001. Measuring Usability with the USE questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG*, 8 (1).
- Marathe, S. and Sundar, S.S., 2011. What Drives Customization? Control Or Identity? In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 781-790.
- McCarthy, J. and Wright, P., 2004. *Technology as Experience*. Cambridge, Massachusetts: MIT Press.
- McLellan, S., Muddimer, A. and Peres S.C., 2012. The effect of experience on System Usability Scale Ratings. *Journal of Usability Studies*, 7 (2), 56-67.
- McLoughlin, M. and Ciolfi, L., 2011. Design Interventions for Open-Air Museums: Applying and Extending the Principles of “Assembly”. In *Proceedings of the 2011 annual conference on Human factors in computing systems (CHI '11)*, 7-12 May 2012 Vancouver. New York: ACM Press, 553-556.
- Moggridge, B., 2007. Interaction Design (IDEO). *Innovation Forum*, Potsdam, Germany. Available from: [http://interface.fh-potsdam.de/innoforum/english/09\\_videos.php](http://interface.fh-potsdam.de/innoforum/english/09_videos.php) [Accessed 19 July 2012]
- Nisbett, R.E. and Wilson, T.D., 1997. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35 (4), 250-256.
- Norman, D.A., 2004. *Emotional design: Why we love (or hate) everyday things*. New York: Basic Books.
- Phau, I. and Lau, K., 2001. Brand personality and consumer self-expression: Single or dual carriageway? *The Journal of Brand Management*, 8 (6), 428-444.
- Quinn, J.M. and Tran, T.Q., 2010. Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived

- usability. In *Proceedings of the 2010 annual conference on Human factors in computing systems (CHI '10)*, 10-15 April 2010 Atlanta. New York: ACM Press, 353-362.
- Raita, E. and Oulasvirta, A., 2011. Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers*, 23 (4), 363-371.
- Rondeau, D.B., 2005. For mobile applications, branding is experience. *Communications of the ACM*, 48 (7), 61-66.
- Sauer, J. and Sonderegger, A., 2010. The influence of product aesthetics and user state in usability testing. *Behaviour & Information Technology*, 30 (6), 787-796.
- Saville, D.J., 1990. Multiple Comparison Procedures: The Practical Solution. *The American Statistician*, 44, 174-180.
- Solomon, M., Bamossy, G., Askegaard, S. and Hogg, M.K., 2010. *Consumer Behaviour: A European Perspective*. Essex: Pearson Education.
- Tossell, C.C., Kortum, P., Shepard, C., Rahmati, A. and Zhong, L., 2012. An empirical analysis of smartphone personalisation: measurement and user variability. *Behaviour & Information Technology*, 31 (10), 995-1010.
- Tractinsky, N., Katz, A.S. and Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers*, 13 (2), 127-145.
- Tullis, T. and Albert, B., 2008. *Measuring the User Experience: collecting, analysing and presenting usability metrics*. Burlington: Morgan Kaufmann.
- Tzou, R.C. and Lu, H.P., 2009. Exploring the emotional, aesthetic, and ergonomic facets of innovative product on fashion technology acceptance model. *Behaviour & Information Technology*, 28 (4), 311-322.
- Van Schaik, P., Hassenzahl, M., and Ling, J., 2012. User experience from an Inference Perspective. *ACM Transactions on Human-Computer Interaction*, 19 (2), 1-25.
- Vermeeren, A.P.O.S., Law, E.L. Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K., 2010. User experience evaluation methods: Current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI '10)*, 16-20 October 2010, Reykjavik, Iceland. New York: ACM Press, 521-530.

## **Figure List**

Figure 1. Perceived hedonic and pragmatic quality of a mobile device and a mobile application in the context of user experience laboratory evaluations.

Figure 2. Test application on Devices A and B.

Figure 3. Measured evaluation constructs.

Figure 4. Average scores and results from two sample T-tests for four evaluation constructs with significant, or close to significant differences.

Figure 5. A proposed classification for the mobile device characteristics that influence the perceived hedonic and pragmatic quality of a mobile device.