# User Experience Evaluation in an Automotive Context

Moritz Körber, Armin Eichinger, Klaus Bengler and Cristina Olaverri-Monreal, *Member IEEE*
Institute of Ergonomics, Department of Mechanical Engineering
Technische Universität München, Germany
Email: (koerber, eichinger, bengler, olaverri)@lfe.mw.tum.de

***Abstract*—Tough competition in the automobile market makes it very important to exceed customer expectations with in-vehicle devices or applications in order to stand out in an already saturated market. Research on human-machine interaction (HMI) showed that user satisfaction is not only influenced by usability factors but also by user experience (UX). Since UX in the automotive context has not yet been comprehensively investigated, this article presents a feasible research method for the measurement of UX. We rely on the fulfillment of psychological needs as a user experience measurement and conducted three online surveys to develop a questionnaire for that purpose. We analyze nine need scales, each with 10 items, and reduced the item count afterwards for a shorter form of the questionnaire. Results reveal that the method successfully measured the intended needs. Future work with respect to a further extension and improvement of the method is discussed.**

## I. INTRODUCTION

In the automobile market, increasing competition makes it very important for car manufacturers to not only meet customer expectations but to exceed them in order to enter a market and to achieve growth rates. A high level of satisfaction and a positive evaluation of their products are needed to gain customers and their loyalty. Research in human-machine Interaction (HMI) has evolved from formulating rigid guidelines and systematic testing, to include methods of user-centered design and context inquiry. Current approaches broaden the discipline's methodological horizon by trying to operationalize affective criteria and UX (UX) more explicitly [1]. Previous research has revealed that in the evaluation of the user satisfaction with a certain product, usability is not the only relevant factor [2]. As most of the car manufacturers already perform usability testing within the quality assurance process, it is important to include additional factors, like UX, in the evaluation that determine if a certain application or in-vehicle device meets or even exceeds users' expectations in order to be the user's device of choice.

To evaluate and compare the UX of their products, car manufacturers need measurement instruments that go beyond typical usability measures and are specified for the purpose of user experience evaluation. Therefore we present in this paper a feasible research method for the quantitative measurement of UX in an automotive context. For this purpose, we developed a questionnaire as a quantitative measurement instrument. We analyzed 90 items by means of three different online surveys and reduced them to 30 items in a second step.

The remainder of this paper is organized as follows: The section II gives an overview of the related work and states our approach on measuring UX. Section III explains the methodology we used to formulate the questionnaire and presents the procedure of the evaluation. In section IV the results of the data analyses are presented. Finally, the last section V concludes the paper with a discussion of use cases and future research topics.

## II. RELATED WORK

### A. Definition of UX

Although previous studies highlight the significance of UX in product interaction and users' product evaluation, besides some shared understandings [3] no consensus about a general definition for it exists. As a result, many studies refer to UX as a fuzzy construct with no stated boundaries associated with a sometimes arbitrary number of constructs like: fun, emotional, hedonic or aesthetic qualities. But both a meaningful definition and a measurement approach of UX have to focus on construct clarification and construct validity to enable the development of valid measurements.

It is for example important to clarify the relationship between user experience and usability. The ISO-Definition of usability (1998) contains the construct user satisfaction, that seems to concern emotions, but in this case it represents perceptions about the interaction with the product in instrumental terms like perceived ease of use. Therefore, it is not appropriate as a synonym for UX. Instead of objective usability criteria, in UX research subjective criteria and the individual experience with the product are important. Hassenzahl [4] agrees with this differentiation and states that people perceive product interaction according to two different qualities: pragmatic and hedonic. Pragmatic interaction is goal oriented, instrumental, and practical. Usually products with a high pragmatic value are work-oriented and are measured by usability criteria. On the contrary, hedonic interaction is related to leisure products which have their focus at least partially on fun and entertainment.

In this study, we follow this distinction and consider usability to be a necessary pre-condition to guarantee UX. Our measurement, therefore, excludes usability and is merely focused on UX. But this distinction is missing a description of the nature of the experience. For this aspect we refer to the ISO definition of UX which is: "a person's perceptions and responses that result from the use or anticipated use

 DOI: 10.1109/IVS.2013.6629440

of a product, system or service" [5]. We are in line with this definition as we interpret UX as a unique, subjective, and individual experience that is evoked in the interaction with a system. Consequently, we focus on a quantitative measurement of UX encompassing the whole subjective experience with the in-vehicle device or application. In the next section, we explain in detail the method to empirically assess the user experience in interaction with a product.

### *B. Relationship between UX and Needs*

In [6] it was suggested that - according to the activity theory [7] - people have three hierarchical levels of goals: on top there are self-referential be-goals which "motivate action and provide it with meaning". On the second level, do-goals represent a concrete outcome of an action one wants to achieve. The lowest level consists of motor-goals, which represent the execution of the task. For instance, "using a lane change assistant" is a do-goal derived from the be-goal security. The motor-goal, in this example, is to recognize and be aware of the system warnings through the use of the senses of each person.

Hassenzahl [8] states that the pragmatic quality and usability of a product only refers to the achievement of do-goals. Additionally, the hedonic quality refers to the achievement of be-goals, meaning that it deals with psychological human needs such as the need for security or relatedness with others. A psychological need is defined by the authors in [9] as innate energizing states or tension, that are related to an individual's psychological well-being. Satisfaction of needs is important to achieve an ongoing sense of integrity, well-being and balance. Deprivation of a need leads to tension and imbalance and thus impels one to action. Therefore the satisfaction of needs has been discussed as a driver for action and experiences for a long time [10], [11]. Consistent with that definition, Hassenzahl [8] argues further that the fulfillment of be-goals, i. e. basic human needs, is the cause for positive emotions and pleasurable experiences in product interaction and leads to the association of hedonic attributes to the product. His assumption is based on a study by Sheldon and colleagues [12], where a connection was established between satisfying life events and the fulfillment of needs in that situation.

His assumption has been supported in [6] as a significant relationship between seven needs (competence, relatedness, popularity, stimulation, meaning, security, and autonomy) and positive affect in interacting with products like mobile phones or computers was found. Based on this research we assume that positive affect in product interaction is mediated by the fulfillment of needs. Whether a need is salient in a situation depends on the user's goal with respect to a situational context. For instance: if one is driving a car and is not sure if the fuel is sufficient to reach a destination, one is in need of security. A product like a fuel gauge fulfills that need and using it leads to positive emotions. In conclusion, we intend to measure the need fulfillment when interacting with a product in a vehicular context as a quality of the UX and the positive affect as the outcome of the interaction.

A crucial characteristic of UX and the salience of needs is their dependency on context which represents the whole environment of the usage situation that is independent from the product itself. With the aid of a story it is possible to embed the interaction with the product into a context. A story is a narration about a specific interaction of a character with a system located in a specific physical and emotional context [13], [14], [15]. Because the story is focused on the character's feelings and goals, a personal identification with him is possible. If the product interaction is embedded in a context, a unique instance of an experience rather than experiencing in general arises. In this study we use stories to create a context for a certain experience with a particular in-vehicle device. They were previously intentionally created to address certain needs that should be fulfilled in the product interaction with the studied in-vehicle devices. Thus, in this study we demonstrate the measurement of the user experience of an existing device combined with an existing story (for this device). As we evaluate vehicular products, our stories are set in an automotive context. We assume that the need fulfillment leads to positive affect and thus a positive experience.

Several works have investigated methods to measure UX, most of them focusing on non-vehicular products. For example, the authors in [16] developed a measurement instrument for perceived website aesthetics, dealing merely with appearance and not with experiences itself. Following the concept of independent hedonic and pragmatic factors the authors in [17] used in their study the questionnaire Attrakdiff2 [18] for measuring hedonic and pragmatic qualities in order to assess the attractiveness of a software product through a semantic differential and hedonic factors covered by three scales, similarly to the method presented in [19]. Both measurement concepts contained usability measures and, therefore, fewer items measured UX itself.

The proposed questionnaire is specifically designed for assessing UX. Instead of a semantic differential we use a Likert-type format as answer mode for nine scales. The reason for this is that a semantic differential deals rather with the personal meaning of a concept or affective associations with it and is therefore not ideal to assess the perception of the experience in terms of needs. We also consider the likert-type format to be more appropriate for this summative evaluation, because of the ephemeral nature of the affective associations a semantic differential is intended to capture and assess.

In an automotive context, the authors in [20] measured UX by investigating the fulfillment of five psychological needs and the positive affect in the interaction with an in-car device to interchange information with other drivers. They used 5 items per need, that they partially created themselves and adopted from [12]. The authors showed the expected fulfillment of needs and obtained a positive affect rating. However, no statistical link in terms of correlation between the fulfillment of needs and the affect was drawn.

Since interaction with a product should lead to a certain need fulfillment, we assume that by measuring this fulfillment we can obtain quantitative data related to the subjective, holistic

UX. For this reason, our questionnaire is based on a similar approach, in which we analyze and revise 10 positively and additionally negatively formulated self-created items for nine instead of five scales. The nine scales are based on the scales used in [12]. By means of a post item analysis we create a shorter form of the questionnaire resulting in 3-4 items for each scale. Therefore we attain a short, but broad measurement of the need fulfillment. We assume a relationship of the need fulfillment with positive experiences and thereby positive affect. The next section describes our approach of measuring need fulfillment.

## III. METHODOLOGY

For each need scale, 10 items were created; 7 positively and 3 negatively formulated. To find out which items reflected the needs selected and to reduce poor performing items, we created three online surveys, two of them based on a different story set in an automotive context and presenting a person interacting with a novel in-vehicle system. The third survey was based on three stories: a dream about driving a new car, two drivers competing with each other, and a reflection of memorable experiences with a car. All presented stories were related to certain needs. The participants in the surveys were instructed to put themselves in the role of the story's main actor and then imagine how it would feel to interact with the device or to drive the car of the person in the third story. For all items, a 5 point scale was used as response mode, ranging from "I disagree completely" to "I agree completely". The tested in-vehicle systems consisted of:

- a device that provided the driver with a visual and haptic feeling about the state of the battery in an electrical vehicle;
- a device to approach points of interest chosen by the vehicle passenger;

To collect the data, the subject was first presented with a story in form of a storyboard with subtitled text and then a questionnaire with the respective items to be filled with the most appropriate answer. We describe in the next lines the needs and stories of the three surveys:

1) The storyboard of the first survey described the holiday trip of a family, characterized by new in-vehicle technologies, that were available to the passengers to interact with. The needs that should be evoked by the story were "relatedness", "stimulation" and "competence".
2) In survey 2, the story showed a man driving home from work, who used a different kind of devices to reach his goals. The needs addressed in this story were "security", "physical form" (i.e. well-being, health status, etc.), and "autonomy".
3) Finally, survey 3 consisted of 3 commercials of the automotive sector. The stories in these commercials addressed the needs "popularity" (first commercial), "competition" (second commercial) and "conservation of meaningful things" (third commercial) (i.e. be reminded of something through a souvenir). In addition to these three needs, we included a short scale for "stimulation", "competence" and "relatedness" for cross-validation purposes to show that the presentation media, video or written text, did not have an effect on the results of the survey.

Figure 1 depicts a section of the stories that the user could visualize.

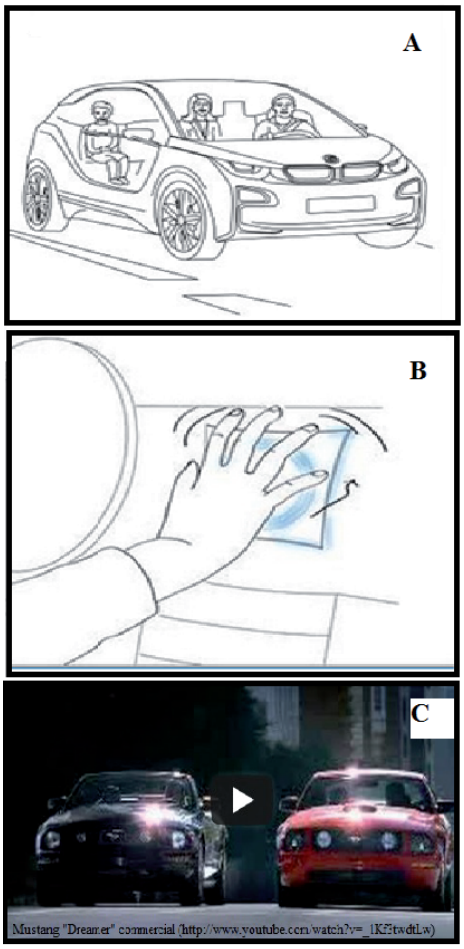


Fig. 1. Study A, B and C illustrating a section of the story that was presented to the subjects.

The items were analyzed descriptively in terms of mean, standard deviation and corrected item-total correlation. Each statistical test was performed with a significance level of $\alpha$ = .05. If the mean of the need was significantly different from the scale mean of 3, we considered this as a salient need, because 3, was the neutral middle category and its selection represented that a certain need was neither fulfilled nor unfulfilled.

For each scale we conducted a t-test as statistical method to test differences in means with metric data to determine if the need was salient for the subject in the situation. Additionally the reliability of each scale was assessed by calculating the internal consistency through Cronbach's $\alpha$. The internal consistency represents the average correlation between the items of a scale and depicts to what extent the items of one scale are in the same relationship to each other.

Additionally, since all items of the scales should measure

a certain need, we conducted confirmatory factor analyses (CFA) for each scale to determine if the data was consistent with the construct's theoretical understanding and hypothesized structure and confirm that each need scale consisted of only one factor. Since three needs where measured in each of the three studies, we also conducted a factor analysis over the respective three needs in each study to confirm that they measured in total nine independent need scales. To reduce the item count, and thus achieve a simplified methodology, three criteria (mean, standard deviation and corrected item-total correlation) were used to evaluate and reduce the number of items where applicable.

As the data revealed little differences in standard deviation and corrected item-total correlation (see IV section), we mainly used the mean value as the main selection criteria, choosing the items with the highest means. This selection process reduced the item count to four items per need scale (three positive, one negative). Again, reliability was measured and if it did not fall below .60 we again reduced the respective scale by one positive item.

## IV. Data Evaluation Results

### A. Sample

Table I describes the sample in the three online surveys. The participation rate in the survey was with a total of 207 responses very high. The highest survey participation rate was reached in the Survey 1, followed by Survey 2 and then Survey 3 (see table I). Since the rate for all three surveys was comparable, the data evaluation was performed based on these numbers. The complete sample consisted of 34.78% females and 65.22% male participants with an average age of 30.09 years.

| | N | Female% | Male% | Mean Age | Age SD |
|---|---|---|---|---|---|
| Survey 1 | 81 | 33.3 | 66.7 | 29.56 | 8.77 |
| Survey 2 | 65 | 32.3 | 67.7 | 29.02 | 9.37 |
| Survey 3 | 61 | 39.3 | 60.7 | 30.77 | 10.36 |

TABLE I
Data sample in the three online surveys

### B. Data Analysis

Table II shows the means and reliability for the need scales (computed as the mean over all items) and the respective t-test results including the degrees of freedom and the effect sizes calculated as the point-biserial correlation coefficient r. The data revealed little differences in standard deviation and corrected item-total correlation. The effect sizes complement the p-value as they state the standardized magnitude of an effect independent from the size of the sample. Because of that it is often taken as a measurement for practical relevance. The effect sizes were in most cases large and except for one need (physical form), all needs were stated as fulfilled in the respective situations even if the mean value is lower as expected.

The reliability of every scale was above the threshold for the at least acceptable reliability of $\alpha$ = .60. This means that the scales are in line with this psychometric standard. As most of the scales achieved a reliability above $\alpha$ = .80, the reliability of the scales can be regarded as very good.

For the purpose of cross-validation we used items from the scales stimulation and competence from survey 1 again in survey 3. This was possible because the stories in survey 3 also addressed these needs. As expected, these two needs were stated as salient and, therefore, the results confirmed that the story-need-relationship is independent from its form of presentation (storyboard, video).

Additionally discriminant validity was assessed on a small scale by using a need (relatedness) for a story, which was not specifically addressing this need. Data revealed that the need was in fact not salient in that case and thus discriminant validity could be confirmed.

Results of the factor analysis in all studies revealed that the negatively formulated items blurred the factor structure and seemed to have an effect on the results because of their different characteristics compared to the positively formulated items. When negative items were left out, in survey 1 and 3, the factor analysis confirmed in each case a three factor model as expected. Also, for each need scale a one factor model was the best fitting model, meaning that the items of one need scale measured in fact the same need.

In survey 2 the factor structure was not as clear as in the survey 1 and 3:

1) First of all, a one factor model resulted only for a one scale need (autonomy). The items did not always measure the intended needs in the other two scales;
2) Second, in survey 2 there was evidence for a three factor model, but 4 items from two need scales were not loading on their respective need scale factor. 3 of them should correlate with the "physical form" factor, which is the only need scale with a mean value lower than the scale mean. As a result, the items from survey 2 did not measure three independent needs as expected.

Figure 2 illustrates through 3 plots for each survey, the number of factors as a result of the statistical analysis. The following abbreviations denote the studied needs:

- Survey 1: r = relatedness, c = competence, s = stimulation;
- Survey 2: s = security, ph = physical form, a = autonomy;
- Survey 3: c = competition, con = conservation of what is meaningful, p = popularity;

Three clusters can be observed in the graphic for each studied need and its corresponding items for survey 1 and 3. The figure for survey 2 shows how some items do not fit for the intended need measurement as they load on a different factor (e.g. s6, a8).

To obtain a shorter version of the questionnaire, we reduced the items for each scale by the criteria mean, standard deviation, factor loading and corrected item-total correlation considering a reliability above .60 as the limit for item reduction. As a result, items could be reduced from 10 to 3 items for six need scales. For the other three need scales

items could be reduced from 10 to 4. Table II illustrates the reliabilities for the short scales, denoted by $\alpha$-short-scale parameter with values that approximately range between .60 and .80. Therefore the need scales can be regarded as acceptable reliable.

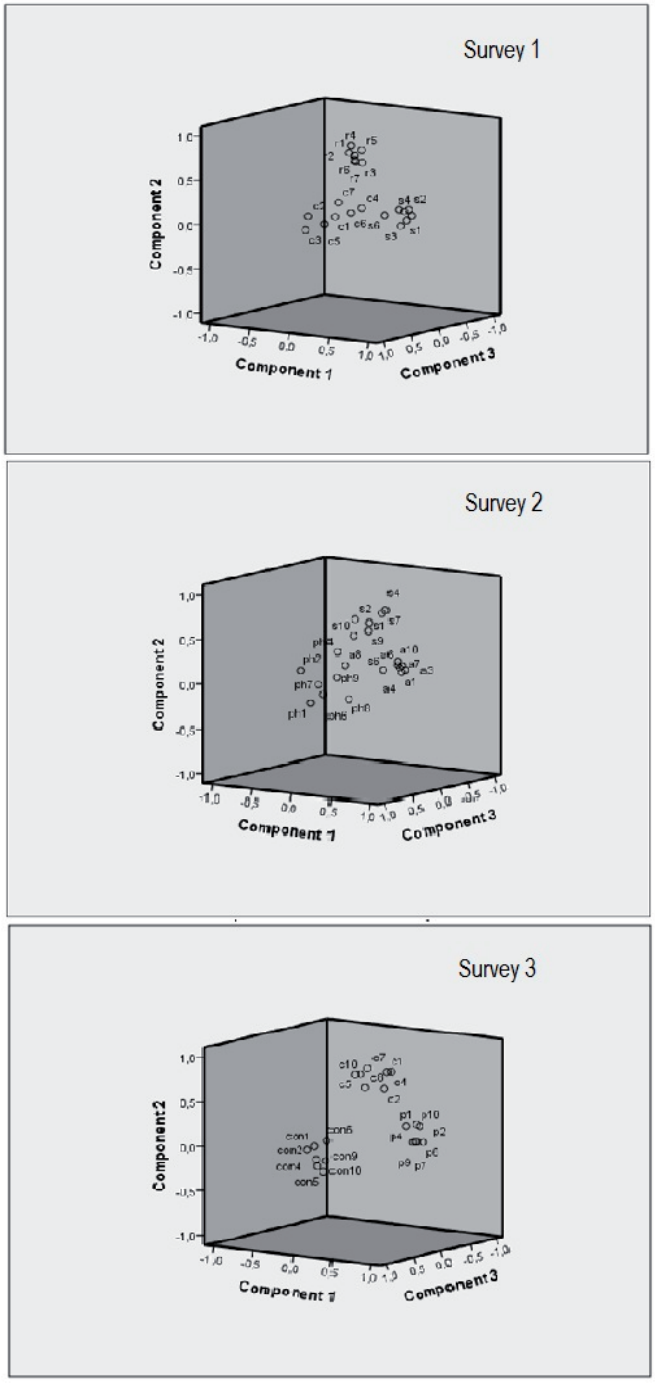


Fig. 2. Factor analysis results represented through a component plot in rotated space for each study and three need scales each.

| Need | M | SD | $\alpha$ | t | df | Pearson's r | $\alpha$-short-scale |
|---|---|---|---|---|---|---|---|
| Relatedness | 3,20 | 0,90 | .92 | 2,02* | 79 | .22 | .79 |
| Relatedness1 | 2.92 | 0.94 | .76 | -.57 | 53 | .09 | |
| Stimulation | 3.37 | 0.81 | .93 | 4.01*** | 78 | .41 | .77 |
| Stimulation1 | 4.03 | 0.65 | .53 | 12.41*** | 60 | .85 | |
| Competence | 3.90 | 0.65 | .89 | 12.40*** | 79 | .85 | .81 |
| Competence1 | 4.03 | 0.72 | .76 | 10.61*** | 55 | .82 | |
| Security | 3.64 | 0.67 | .86 | 7.59*** | 63 | .69 | .74 |
| Physical-form | 2.89 | 70.63 | .84 | -1.40 | 61 | .18 | .70 |
| Autonomy | 3.51 | 0.71 | .87 | 5.62*** | 61 | .58 | .68 |
| Popularity | 3.24 | 0.83 | .88 | 2.22* | 60 | .28 | .73 |
| Conservation of what is meaningful | 3.65 | 0.71 | .86 | 6.75*** | 53 | .68 | .59 |
| Competition | 3.5 | 0.62 | .77 | 6.08*** | 55 | .63 | .69 |

TABLE II
RESULTS OF THE DESCRIPTIVE ANALYSIS IN TERMS OF MEAN, STANDARD DEVIATION AND CORRELATION VALUES AND EFFECT SIZE MEASURED IN PEARSON'S R. (* = P $<$ .05; ** = P $<$ .01; *** = P $<$ .001). THE INTERNAL CONSISTENCY VALUE IS DENOTED BY $\alpha$

## V. CONCLUSION AND FUTURE WORK

Aim of this study was to develop a questionnaire for need fulfillment in product interaction. We created 10 items for each need scale and conducted three online surveys to analyze these items. We can state that although the mean value resulting from the analysis was small, the developed questionnaire successfully measured the anticipated need fulfillment in the story situations and effect sizes were in most parts large. The small value could be due to the research design, since the stories chosen were not especially designed for certain needs, but to provide the user with a background about a certain in-vehicle application.

Since the design of the questionnaire items is not specific of a certain product type and the items are rather abstract, we can conclude that the questionnaire is a reliable instrument to assess the UX based on psychological needs for a variability of product types in an automotive context.

As already stated in the introduction, the questionnaire is not assessing usability, thus we could achieve that 30 items reflected the whole experience with an in-vehicle device in relation to needs. This tool requires less than 10 minutes processing time for the participants; therefore we have achieved a low time consuming short but powerful measurement tool for UX. Additionally, this tool makes it possible to just pick out relevant scales for usage reducing even more the time for the user. Future goals are to pursue with the research in terms of validation and revise and confirm the factors structure. Furthermore, convergent and discriminant validity has to be assessed through further existing questionnaires [12]. Additionally, changes in the participant's affect as outcomes of the need fulfillment could be measured by the Positive and Negative Affect Schedule PANAS [21] or it's short form [22]. In this study we only examined the anticipated use of product concepts. This questionnaire must still be validated by real interaction with in-vehicle devices. To extend the theoretical fundament for this research, future studies could analyze the relationship between needs and emotions in greater detail. Cognitive processes that explain why the fulfillment of a need is connected with positive experiences also have to be revealed.

## VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the "CAR@TUM User Experience" project by BMW AG and the valuable contribution to the project of Marc Landau (Institute of Industrial Design, Technische Universität München), through his stories sketches.

## REFERENCES

[1] S. Bødker, "When second wave hci meets third wave challenges," in *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. ACM, 2006, pp. 1–8.

[2] A. Sonderegger and J. Sauer, "The influence of design aesthetics in usability testing: Effects on user performance and perceived usability," *Applied Ergonomics*, vol. 41, no. 3, pp. 403–410, 2010.

[3] V. Roto, H. Rantavuo, and K. Väänänen-Vainio-Mattila, "Evaluating user experience of early product concepts," in *Proceeding of the International Conference on Designing Pleasurable Products and Interfaces DPPI09*. Citeseer, 2009, pp. 1–10.

[4] M. Hassenzahl, "Experience design: Technology for all the right reasons," *Synthesis Lectures on Human-Centered Informatics*, vol. 3, no. 1, pp. 1–95, 2010.

[5] I. DIS, "9241-210: 2008. ergonomics of human system interaction-part 210: Human-centred design for interactive systems (formerly known as 13407)," *International Organization for Standardization (ISO). Switzerland*, 2010.

[6] M. Hassenzahl, S. Diefenbach, and A. Göritz, "Needs, affect, and interactive products–facets of user experience," *Interacting with Computers*, vol. 22, no. 5, pp. 353–362, 2010.

[7] V. Kaptelinin and B. Nardi, *Acting with technology*. Mit Press, 2006.

[8] M. Hassenzahl, "User experience (ux): towards an experiential perspective on product quality," in *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine*. ACM, 2008, pp. 11–15.

[9] R. Ryan and E. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American psychologist*, vol. 55, no. 1, p. 68, 2000.

[10] A. Maslow, R. Frager, and J. Fadiman, *Motivation and personality*. Harper & Row New York, 1970, vol. 2.

[11] H. Reis, K. Sheldon, S. Gable, J. Roscoe, and R. Ryan, "Daily well-being: The role of autonomy, competence, and relatedness," *Personality and Social Psychology Bulletin*, vol. 26, no. 4, pp. 419–435, 2000.

[12] K. Sheldon, A. Elliot, Y. Kim, and T. Kasser, "What is satisfying about satisfying events? testing 10 candidate psychological needs." *Journal of personality and social psychology*, vol. 80, no. 2, p. 325, 2001.

[13] T. Erickson, "Notes on design practice: Stories and prototypes as catalysts for communication," 1995.

[14] W. Quesenbery and K. Brooks, *Storytelling for user experience: Crafting stories for better design*. Rosenfeld Media, 2010.

[15] P. Björndal, M. Rissanen, and S. Murphy, "Lessons learned from using personas and scenarios for requirements specification of next-generation industrial robots," *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, pp. 378–387, 2011.

[16] T. Lavie and N. Tractinsky, "Assessing dimensions of perceived visual aesthetics of web sites," *International journal of human-computer studies*, vol. 60, no. 3, pp. 269–298, 2004.

[17] J. Isleifsdottir and M. Larusdottir, "Measuring the user experience of a task oriented software," in *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island*. Citeseer, 2008, pp. 97–102.

[18] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Mensch & computer*, vol. 2003, 2003, pp. 187–196.

[19] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," *HCI and Usability for Education and Work*, pp. 63–76, 2008.

[20] M. Knobel, M. Hassenzahl, M. Lamara, T. Sattler, J. Schumann, K. Eckoldt, and A. Butz, "Clique trip: feeling related in different cars," in *Proceedings of the Designing Interactive Systems Conference*. ACM, 2012, pp. 29–37.

[21] D. Watson, L. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales." *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.

[22] K. Kercher, "Assessing subjective well-being in the old-old the panas as a measure of orthogonal dimensions of positive and negative affect," *Research on Aging*, vol. 14, no. 2, pp. 131–168, 1992.