Measuring user experience in a railway related environment

Birgit MILIUS

Technische Universität Braunschweig, Germany

Abstract. Railway research has shifted from looking at replacing humans to integrating them. Often, usability was used to look at the integration process. In recent years, this concept was criticized due to its strong focus on task- and goal related effectiveness and efficiency. A new concept called user experience (UX) looks at emotions and the perception of non-instrumental qualities. In this paper, we look at how to measure UX and what needs to be taken into account to derive consistent and meaningful results.

Keywords. User experience, usability, driving simulator, safety

1. Once upon a time…

At the Technische Universität Braunschweig a study was performed to learn more about the acceptance of virtual reality devices. In short, an Oculus Rift DK2 with leap motion was used in an experiment with students. The students had to perform two tasks. First, they had to learn to catch and stack blocks. Afterwards, in a second simulation they were train control operators in a virtual reality environment and had to set a route for an incoming train. The experiments are described in detail in Burkhardt et al. (2017). Several different methods were used to derive a feedback from the students. As we expected that with a rather new technology curiosity and fun were probably more or at least as relevant to the students as usability, we used UX questionnaires to gain feedback. To also learn about the UX methodology, we used two different, German language questionnaires which differed slightly in length and presentation: Attrakdiff and UEQ.

The results were surprising: Both questionnaires argue that they measure pragmatic (PQ) and hedonic (HQ) quality as well as attractiveness (ATT) on a scale from -3 to +3, with results usually are between -2 and +2. The results of the questionnaires based on the evaluations of 20 participants differed very much from each other (table 1). How this possible, where was the mistake?

Table 1: Results of the UX questionnaires for the experiment "Virtual Reality (VR)"

	Attrakdiff	UEQ
PQ	0,66	1,25
HQ	0,81	1,88
ATT	1,11	1,25

2. What is User Experience?

For a long time, development of products focused on usability aspects. Usability is per definition "The extent to which a product can be used by specified users to achieve specified goals with

effectiveness, efficiency, and satisfaction in a specified context of use." [ISO] However, over the last 20 years, some criticism of the concept evolved Hassenzahl et al. (2000). This was mainly based on the fact that satisfaction is part of the usability concept, but not as a property of itself but rather as the result of effective and efficient working.

In the 1990s, several researchers started to consider pleasure and satisfaction as a separate research area. By then, not even the term of user experience was agreed on. They called this research area, e.g. emotional usability Logan (1994) or pleasure Jordan (2000). Other words used to describe emotional aspects of usability were/are joy of use, aesthetics or emotions

UX is still a rather young area of research and many aspects, e.g. definitions are not finally agreed on. Examples for definitions are e.g.

- All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they are using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it. (Alben (1996))

- [UX encompasses] all aspects of the end-user's interaction with the company, its services, and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer, without fuss or bother. Next come simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. (User Experience (Nielsen-Norman Group (2007)))

- [UX is] a result of motivated action in a certain context. (Mäkelä & Fulton Suri (2001))

- [UX is] a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.). (Hassenzahl & Tractinsky (2006))

Hassenzahl, Law and Hvannberg (2006) emphasize three main aspects where UX is going beyond the traditional usability metrics: It is holistic, looks in detail at the subjective opinions of people and stresses the importance of positive outcomes of technology use or possession. Hassenzahl et al., (2000) proposes a theoretical model, which was analyzed and proven in several studies, which differentiates between the pragmatic quality (instrumental qualities that are qualities to reach a certain goal) and the hedonic quality[1] (non-instrumental qualities). Hedonic aspects are further distinguished as stimulation, identification and evocation (Hassenzahl (2003)). The model further introduces attractiveness as a result of both pragmatic and hedonic quality (Hassenzahl et al. (2000), Hassenzahl (2002)). Newer research has questioned this and has assumed that this relation changes depending on the evaluated system, especially in a business setting where one could expect that pragmatic aspects dominate. However, this hypothesis could not be proven (Schrepp, Held, Laugwitz (2006)).

3. On the importance of UX for industrial applications
At first it might be surprising that the actual experience – looking at emotions like fun and

---

[1] Hassenzahl (2001) was the first to introduce the term hedonic quality for the emotions and affects experienced by using a product, using an artificially created term rather the correct term of hedonistic to avoid associations which were otherwise to be expected.

happiness - should be taken into account in a business setting. But research has shown that a positive experience while working will lead to a more effective, efficient and reliable employee. As an example, Schrepp, Held and Laugwitz (2006) show that hedonic quality has also an important impact on the attractiveness of business management software that is used for work purposes. In Hassenzahl (2003) and Vorderer (2005), the authors make it clear that analogies exist between gaming and working. In both cases tasks have to be performed regularly. Some tasks might have differing goals. Difficulties have to be solved and the person has to develop knowledge. If it is possible to transfer the motivation and experiences often associated with gaming into the work area, this can help to solve complex tasks, and to immerse one more into the work problems. Also, Hassenzahl, Burmester and Koller (2003) show that hedonic quality of a product can lead to a higher rating of the overall attractiveness of that product. Igbaria et al. (1994) has find that user acceptance of a new technology is affected both by perceived usefulness and perceived fun. As user acceptance is a very important factor when introducing new systems, the fun – that is the hedonic quality of a business product – should be taken into account.

Also for business reasons, taking user experience into account can be sensible. Often in a business context, for the same functionality several different systems exist. Here, joy of use can be a mean of distinguishing one product from another.

4. Measuring UX

To take user experience into account when developing systems for a business situation and use it as a design goal, it is necessary to be able to actually measure it. This is a difficult, as several different approaches and influencing aspects can be distinguished. As a researcher, one usually wants objective results. However, emotions are very difficult to be picked up in an objective way. When UX is measured subjectively, it becomes more difficult to compare as, e.g. personal experiences and preferences will influence the result. Also, emotions exist only a short time and can change quickly and often over the time of an experiment. Using, e.g. questionnaires after an experiment will probably only gather some aspects, but not all. We also have to remember, that UX results from a combination of the user, the product and the situation the product is used in. This means that, e.g. the setting for an experiment might influence significantly how the UX is rated. Even though the problems continue to exist and will influence the results, questionnaires are still the easiest way to measure UX and several ones are available.

5. Attrakdiff and UEQ

The authors are aware that other questionnaires for measuring UX do exist. The ones presented here were chosen because they seemed comparable and easy to use. Also, lots of literature explaining the methods as well as discussing applications exist so that an in-depth analysis seemed possible.

5.1 What is Attrakdiff?

Attrakdiff is based on the UX model of Hassenzahl as explained above. Hassenzahl refers to pragmatic and hedonic qualities as well as attractiveness. Evocation is part of the model but not part of the questionnaire. Interestingly, while in the paper attractiveness is explained as being the result of hedonic and pragmatic quality, in the assessment the overall attractiveness is not calculated but evaluated using adjectives. In our opinion, this is sensible as otherwise a fixed relation between pragmatic and hedonic quality would have to be assumed. In our opinion that is something one cannot be completely sure about.

The construction process of Attrakdiff is described in Hassenzahl, Burmester and Koller (2003). The first application was for the evaluation of a website. The results were the input in a factor analysis. The adjectives used in the final questionnaire can be seen in table 3.

Attrakdiff is presented online and should be used online. The results are immediately calculated and presented.

5.2 What is UEQ?

The construction process of UEQ is described in Laugwitz, Held and Schrepp (2008). This resulted in the construction of a 26 item questionnaire called User Experience Questionnaire (UEQ). The UEQ contains factors to evaluate six qualities (Schrepp (2015)): Attractiveness: Overall impression of the product. Do users like or dislike the product? ; Perspicuity: Is it easy to get familiar with the product? Is it easy to learn how to use the product? Efficiency: Can users solve their tasks without unnecessary effort?; Dependability: Does the user feel in control of the interaction?; Stimulation: Is it exciting and motivating to use the product?; Novelty: Is the product innovative and creative? Does the product catch the interest of users?

Attractiveness is a pure valence dimension. Perspicuity, Efficiency and Dependability are pragmatic quality aspects (goal-directed), while Stimulation and Novelty are hedonic quality aspects (not goal-directed). The adjectives used in the questionnaire can be seen in table 3 of this paper. The questionnaire is presented on paper. The data can then be put in an excel-table which also calculates and presents the results. The UEQ questionnaire exists in English as well as German.

6. Case studies

The case study, which started our in-depth research, was described in the first chapter. To get more experience, a second case study was performed. Every year, students studying the subject of railway systems engineering have to perform several session in the virtual railway lab. They all come with little prior knowledge. After 90 minutes in the lab, performing several tasks typical for train control operators, they were asked to fill out the Attrakdiff and UEQ questionnaires. Unfortunately, not all students showed up and finished the questionnaire. The results from 15 participants are shown in table 2. Again, the results differed but not as strongly as in case study "Virtual Reality (VR)"

Table 2: Results of the UX questionnaires for the experiment "Laboratory"

	Attrakdiff	UEQ
PQ	0,27	0,41
HQ	0,54	0,58
ATT	0,82	1,11

7. Discussion of the results

7.1 Comparison of the model

Both questionnaires are supposed to measure pragmatic and hedonic quality as well as attractiveness. Looking more into detail, the developers argue that hedonic quality consist of different aspects. While Attrakdiff looks at identity and stimulation, UEQ looks at novelty and stimulation. If such a difference does also exist for pragmatic quality cannot be assessed, as Attrakdiff does not give more details. However, as the pragmatic quality relates closely to

usability and as such is well known and often researched, we can assume that the definition of pragmatic quality does not vary so much.

7.2 Looking at the adjectives used to describe the aspects

A detailed comparison of the adjectives is difficult. First of all, German language is varied, so the same emotions can be described with slightly different word. When translating into English, these differences might disappear, but others might arise. The following comparison is based on the German adjectives, as the questionnaires are originally in German.

Regarding attractiveness, there are seven pairs of adjectives in Attrakdiff and six in UEQ. Five pairs are identical. The meaning of the other pairs is clear and they seem to assess the same opinions/emotions, just using different wording.

While UEQ uses 12 pairs of adjectives for pragmatic quality, Attrakdiff uses just seven. Four pairs are identical. Interestingly, the UEQ uses one adjective in two pairs what makes the comparison more complicated. Only UEQ structures the adjectives into the three groups efficiency, perspicuity and dependability. To better understand if the resulting measure of Attrakdiff can be the same as in UEQ, the four pairs of adjectives of Attrakdiff not directly comparable will be grouped into the three groups of pragmatic quality as used by UEQ.

*Table 3: UEQ and Attrakdiff in English, Attrakdiff translations were taken from Hassenzahl, (2004). English verbs with an * were translated by the author; cursive: identical pairs*

	UEQ		Attrakdiff	
ATT	annoying	enjoyable	*repudiative	*inviting
			*discouraging	*motivating
	good	*bad*	*bad*	*good*
	unlikable	*pleasing*	*unlikable	*pleasing
	unpleasant	*pleasant*	*unpleasant	*pleasant
	friendly	*unfriendly*	*friendly	*unfriendly
	attractive	*unattractive*	ugly	beautiful
PQ-E efficiency	fast	slow		
	inefficient	efficient		
	impractical	practical		
	organized	cluttered		
			impractical	practical
			cumbersome	direct
PQ-P perspicuity	not understandable	understandable		
	easy to learn	difficult to learn		
	complicated	*easy*	*complicated*	*easy*
	clear	*confusing*	*confusing*	*clear*
PQ-D dependability	*unpredictable*	*predictable*	*unpredictable*	*predictable*
	obstructive	supportive		
	secure	*not secure*	*unpredictable*	*predictable*
	meets expectations	does not meet expectations		

			unruly	manageable
PQ- ?	technical - human			
HQ-I Identity			isolating	integrating
			amateurish	professional
			gaudy	classy
			cheap	valuable
			Non-inclusive	inclusive
			takes me distant from people	brings me closer to people
			unpresentable	presentable
HQ-S Stimulation	Valuable	inferior	typical	original
	Boring	exciting	*standard*	*creative*
	not interesting	interesting	cautious	courageous
	motivating	demotivating	conservative	innovative
			lame	exciting
			easy	challenging
			commonplace	new
HQ-N Novelty	*Creative*	*dull*		
	Inventive	conventional		
	Usual	leading edge		
	Conservative	innovative		

As you can see in table 3, for all but one pair (technical – human) such analysis was possible. In the author's opinion, the pair technical-human does not fit any of the three categories.

The structure of the adjectives of hedonic quality is the most difficult to analyze. UEQ provides eight pairs in the categories stimulation and novelty, whereas Attrakdiff provides 14 pairs in the categories identity and stimulation. There is no pair which is identical in both questionnaires. Sometimes, the wording is close, e.g. standard/dull vs. creative. Interestingly, whereas UEQ groups this under novelty, in Attrakdiff it is grouped under stimulation.

Looking at the adjective as a German native speaker, it is obvious, that the categorization done in UEQ is comprehensible. The pairs of adjectives describe the same emotions and qualities using different wordings. They are generally easy to understand. The adjectives of Attrakdiff are more difficult to understand and harder to apply. This impression was also obtained by the author based on informal remarks when the students filled out the questionnaires.

Looking at the identity-category, it is easy to see the connection between the adjectives and the label "identity". Looking at the pairs of adjectives in the group stimulation of Attrakdiff, it is obvious that here adjectives regarding stimulation (lame, exciting, challenging, easy…) are mixed with adjectives of the category novelty (conservative, innovative, original, typical…).

In conclusion, both questionnaires look at stimulation and novelty but the Attrakdiff questionnaire does also consider aspects of identity.

7.3 Comparison of the results of the case studies

In this chapter we have a closer look at the results from the case study. We will try to explain where the differences come from and what that means for future UX evaluations. The following

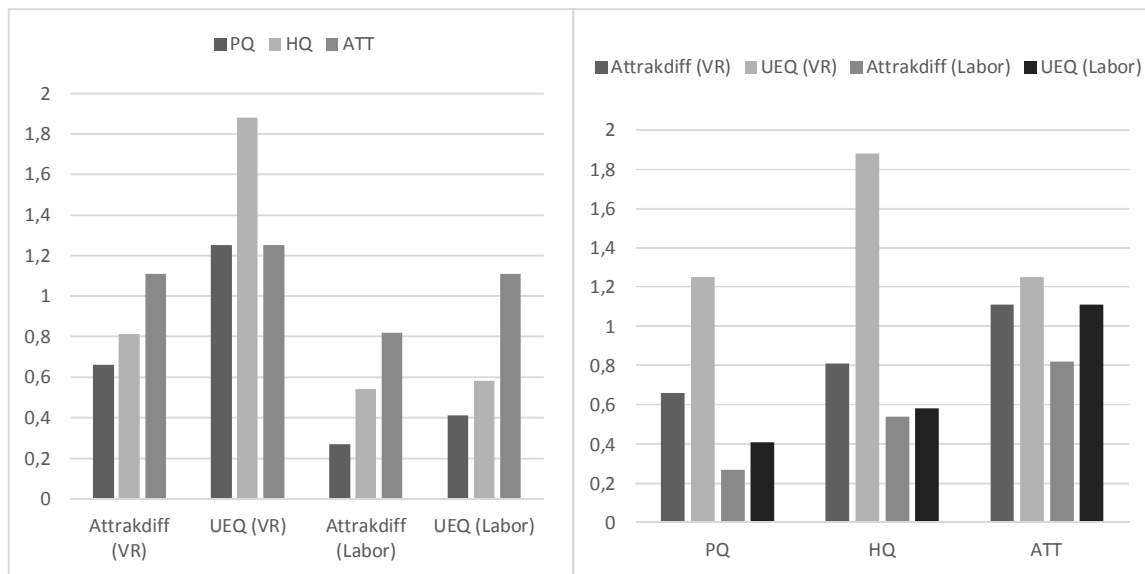statements can be derived from figure 1.

Figure 1: left: Comparison of the results for the different questionnaires and the different case studies; Right: Comparison of the results for the three different aspects Pragmatic Quality (PQ), Hedonic Quality (HQ) and Attractiveness (ATT)

1. In all cases the hedonic quality is rated higher than the pragmatic quality.
2. The pragmatic quality of the interlocking simulation is rated lower than the pragmatic quality of the VR setting.
3. In three of four assessments attractiveness rated higher than anything else, only the UEQ for the VR setting varies from this.
4. For the VR setting, UEQ provides much better results than Attrakdiff. This difference is smaller for the Laboratory-setting.
5. For the Laboratory setting, both Attrakdiff and UEQ measure about the same effects.

Result 1

For the VR setting, the difference is plausible. The fun factor was much more important than the usability aspect. The students came with the wish to experience VR and did not have any specific goal. Interestingly, also for the more scientific setting the hedonic aspect was higher even though the students were supposed to learn. This could mean that the used simulation software is either very difficult/too complicated, that the students were not well enough prepared or that the students actually enjoyed "gaming" with the simulation.

Result 2

It sheds an interesting light on the used interlocking simulation when the pragmatic quality of a game is rated higher than the pragmatic quality of actual work software. However, this might also result from the fact that pragmatic and hedonic qualities are not independent from each other. In Harbich & Auer (2005) this effect was researched. The authors came to the conclusion that hedonic quality and usability correlate to a big extent. When hedonic quality is rated high, usability ($\approx$ pragmatic quality) is also rated high or higher than expected.

Result 3, 4 & 5

To better understand the difference in the VR evaluation, we look at the results qualitatively and quantitatively. In general, the students did not report more difficulties with the questionnaires in the VR setting compared to the Laboratory-setting. In both situations, they did have more problems interpreting the adjectives in Attrakdiff compared to the adjectives in UEQ. We assume that the difference does not result from a general problem with either questionnaire.

To see if single adjectives were responsible for the difference, we calculated the standard deviation for each category (table 4). The results did not show an obvious difference for the VR setting.

Table 4: Average and standard deviation for the categories of UEQ in the VR setting

	Attractiveness	Pragmatic Quality			Hedonic Quality	
		Perspicuity	Efficiency	Dependability	Stimulation	Novelty
mean	1,25	1,31	0,58	0,89	1,71	2,05
standard deviation	0,97	0,74	0,82	0,73	0,75	0,75

Next, we looked at the adjectives for hedonic quality of both UEQ and Attrakdiff. The UEQ results (figure 2) show that all adjectives get a better assessment in the VR setting than in the Laboratory setting. This actually makes sense as the hedonic quality of the VR is much higher than the hedonic quality of the laboratory software. Looking at the results from Attrakdiff (figure 3), we see that two adjectives have a negative connotation. Looking at the mean and standard deviation (table 5), we see that the standard deviation of hedonic quality – Identity (where the adjectives belong to) is very different from the standard deviation of other items. These two adjectives significantly influence the final result. Adjectives with the same meaning are not part of the UEQ questionnaire.
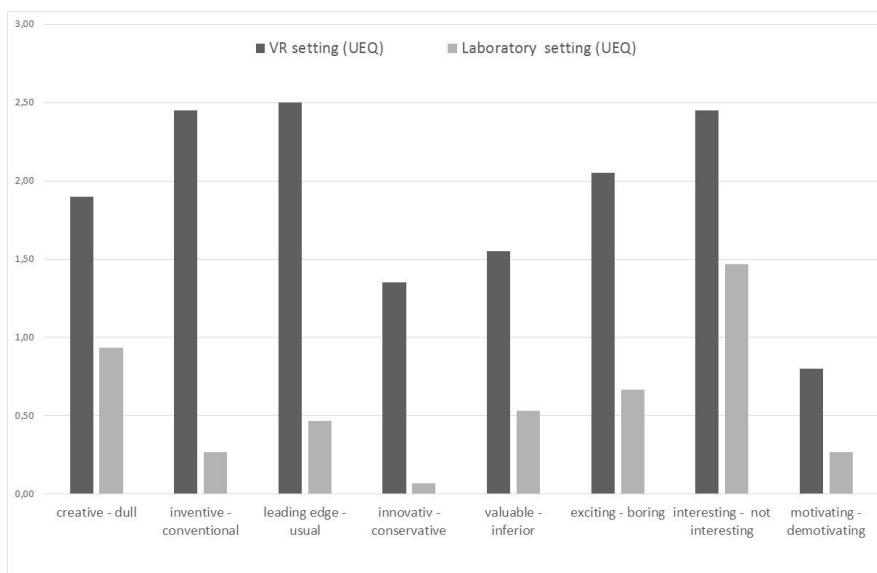

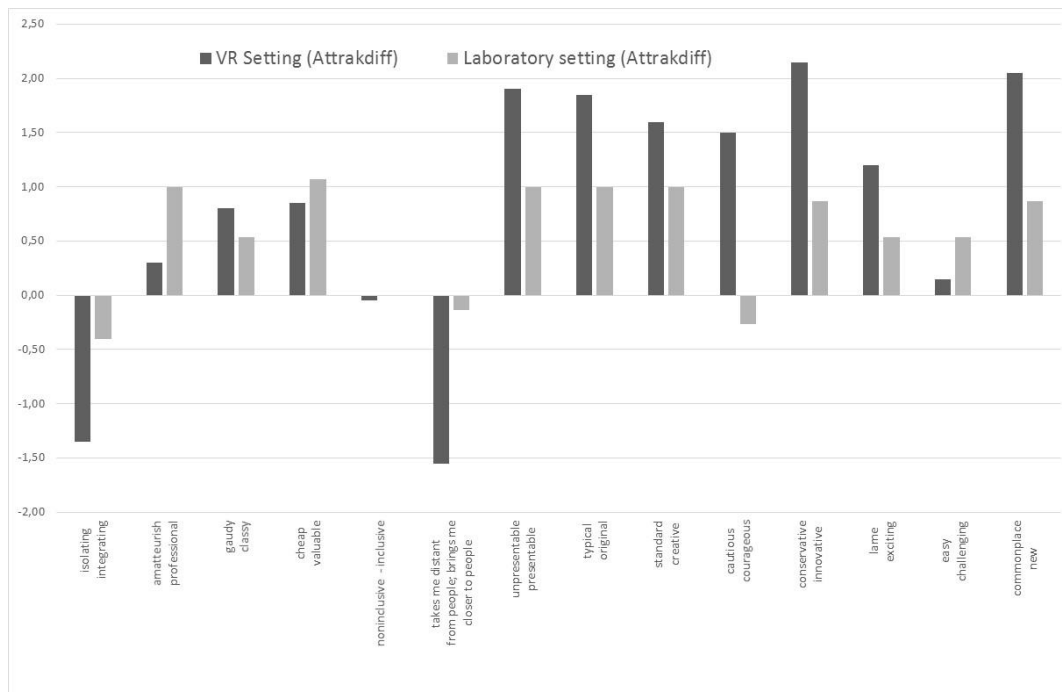*Figure 2: Results for the adjectives of hedonic quality for scenarios VR and Laboratory (UEQ)*

Figure 3: Results for the adjectives of hedonic quality for scenarios VR and Laboratory (Attrakdiff)

Table 5: Average and standard deviation for the categories of Attrakdiff in the VR setting

	Attractiveness	**Pragmatic Quality**	**Hedonic Quality Identity**	**Hedonic Quality Stimulation**
mean	1,11	0,66	0,13	1,50
standard deviation	0,40	0,54	1,14	0,63

8. Discussion of results

In the paper, we could proof that the differences in the result are not due to "errors" but result from the different adjectives used in the questionnaires. It clearly shows that by choosing a questionnaire the final assessment of the product can be strongly influenced and therefore it is necessary to choose the questionnaire wisely. It is necessary to have a closer look at which questionnaires are used for the assessment of UX. We suggest building a database to better understand how the questionnaires work. This could mean to use always two UX questionnaires. This way, it is possible to collect information how they react when used with, e.g. innovative, fun or very pragmatic products.

Measuring UX can give new insight when performing experiments. Especially the relationship and the dependences of hedonic and pragmatic quality should be taken into account when assessing, e.g. the usability of products. Also, it was shown that UX can produce meaningful results for more technical (business) applications and we expect that a significant benefit can be gained from its integration in the evaluation of railway related software, e.g. for train control operators or maintenance.

UX is an exciting area of research and offers lots of interesting, new topics. It is not completely understood yet therefore the results of such projects can significantly influence the research area.

References

Alben, L. (1996). Quality of Experience. *Interactions , 3* (3), 11-15.

Burkhardt M, Christ, A., Dose, D., Grabe, E., Ruth, N. (2017). User TestS of Innovative interaction with a "3D OPERATOR Workstation". In Proceedings of Rail Human Factors Conference 2017, London.

DIN EN ISO 9241: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze (ISO 9241-11:1998). Januar 1999.

Harbich S., Auer S. (2005). Rater Bias: The Influence of Hedonic Quality on Usability Questionnaires. In Costabile M.F., Paternò F. (eds) Human-Computer Interaction - INTERACT 2005. INTERACT 2005. Lecture Notes in Computer Science, vol 3585. Berlin, Heidelberg:Springer.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität In: Ziegler, J. & Szwillus, G. (Hrsg.), Mensch & Computer 2003. Interaktion in Bewegung (pp. 187-196), Stuttgart, Leipzig: B.G. Teubner.

Hassenzahl, M. (2003). Attraktive Software - Was Gestalter von Computerspielen lernen können. In J. Machate & M. Burmester (Hrsg.), User Interface Tuning. Benutzungsschnittstellen menschlich gestalten (pp. 27-45). Frankfurt a. M.: Software & Support Verlag.

Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13, 479–497.

Hassenzahl, M., Platz, A., Burmester, M., Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. In Turner, T., Szwillus, G. (Eds.), Proceedings of the CHI 2000 Conference on Human Factors in Computing (pp. 201–208). New York: ACM, Addison-Wesley.

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. Human-Computer Interaction , 19 (4), 319-349.

Hassenzahl, M., Law, E. L.-C., & Hvannberg, E. T. (2006). User Experience – Towards a unified view. In proceedings of NorciCHI-The Fourth Nordic Conference on Human-Computer Interaction (pp. 1-3). Oslo.

Hassenzahl, M., & Tractinsky, N. (2006). User Experience – a Research Agenda. Behaviour and Information Technology , *25* (2), 91-97.

Igbaria, M., Schiffman, S.J., Wieckowski, T.J. (1994). The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. Behaviour & Information Technology 13, 349–361.

Jordan, P. W. (2000). Designing pleasurable products. An introduction to the new human factors. London, New York: Taylor & Francis.

Laugwitz, B., Schrepp, M. & Held, T. (2008). Construction and evaluation of a user experience questionnaire. In Holzinger, A. (Ed.): USAB 2008, LNCS 5298, pp. 63-76.

Logan, R. J. (1994). Behavioral and emotional usability: Thomson consumer electronics. In M. E. Wiklund, Usability in practice: How companies develop user friendly products (pp. 59-82). Boston: Academic Press.

Mäkelä, A. & Fulton Suri, J. (2001). Supporting Users' Creativity: Design to Induce Pleasurable Experiences. In M. Helander, & H. M. Khalid (Ed.), Proceedings of the Conference on Affective Human Factors. London: Asean Academic Press.

Schrepp, M., Held, T., Laugwitz, B., (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers, 18, 1055–1069.

Schrepp, M. (2015). User Experience Questionnaire Handbook. http://www.ueq-online.org/.

User Experience (Nielsen-Norman Group) (2007). Retrieved June 28, 2017, from Nielsen-Norman Group: http://www.nngroup.com/about/userexperience.html

Vorderer, P. (2005). Ernste Spiele. Retrieved June 30, 2017, from http://www.bpb.de/themen/8GJ1M3,0,0,Ernste_Spiele.html