A Comparison of UX Questionnaires

What is their underlying concept of user experience?

Martin Schrepp
Product Experience CP
SAP SE
Walldorf Germany
martin.schrepp@sap.com

ABSTRACT

To choose the best fitting UX questionnaire for a concrete evaluation project is far from being easy. There is a huge number of different UX questionnaires available. Each of them measures by its scales and items a subset of what we understand by the ill-defined concept of user experience. We analyze a sample of 40 established UX questionnaires and try to work out their differences and similarities. This analysis shows the heterogeneity of the field. In addition, we analyze how the current practice to develop new UX questionnaires adds to this inhomogeneity and lack of common understanding what we measure when we measure UX. Hopefully, this is a first step towards the development of a common framework that helps UX professionals to find their way through the jungle of available UX questionnaires and measurement concepts.

CCS CONCEPTS

Human-centered computing → HCI design and evaluation methods

KEYWORDS

User Experience, Questionnaires, Usability, Measurement

1 Introduction

If a UX professional wants to measure the usability or user experience (short UX) of a product, then he or she has the choice between a huge variety of existing questionnaires. These questionnaires differ in the number and format of the items, the UX quality aspects they measure with their scales, the general approach to measure UX and the product categories for which they are designed.

This makes it difficult to find a suitable questionnaire that can address the research goals and fits in addition to the constraints concerning maximal number of items or maximal response time associated with each real research situation.

We analyze a larger sample of existing UX questionnaires concerning the semantic UX aspects measured by their scales and items.

2 Construction of a sample of questionnaires

For this paper we restrict our investigation to questionnaires that concentrate on the measurement of usability or user experience and for which the items are available without charges in English or German. A literature search resulted in a list of 40 questionnaires that fulfilled these criteria.

The following questionnaires were included in the investigation: Attrakdiff2 [1], AttrakWork [2], CSUQ [3], DEEP [4], e4 [5], EUCS [6], HARUS [7], HED/UT [8], INTUI [9], ISOMETRICS [10], ISONORM [11], meCUE [12], MSPRC [13], NRL [14], PSSUQ [15], PUEU [16], PUTQ [17], QUESI [18], QUIS [19], SASSI [20], SUISQ [21], SUMI [22], SUPR-Q [23], SUS [24], UEQ [25], UEQ+ [26], UES [27], UFOS [28], UMUX [29], Upscale [30], USE [31], UXNFQ [32], VISAWI [33], WAMMI [34], Web-Clic [35], WEBLEI [36], WEBQUAL [37], WEBUSE [38], WEQ [39], and WOOS [40].

Together these 40 questionnaires contain 1248 single items. The questionnaires show huge differences concerning the number of different scales, i.e. the number of different UX aspects that they propose to measure. Some, for example SUS or UMUX, create just one single scale value representing overall UX quality. Others offer many scales, for example WEBQUAL with 9 scales or PUTQ with 8 scales. Some questionnaires follow a modular approach, i.e. they offer a larger set of scales, but do not assume that all scales are used in a single evaluation, for example meCUE with 9 scales or UEQ+ with currently 16 scales.

This list of available UX questionnaires is of course not complete, but contains at least the currently most prominent ones.

3 What do we mean with user experience?

The items of the questionnaires describe the subjective impression of users towards desirable or undesirable properties of products. We use a simple word cloud to give a rough impression concerning the variety of such UX related properties.

The items of the questionnaires from our list were reduced to the used attributes (all other words are removed). Item for which this was not possible were ignored. Then the attributes were unified concerning spelling and their frequency was counted.

https://doi.org/10.18420/muc2020-ws105-236

The following word cloud (generated with the free word cloud generator www.wortwolken.com) shows the word cloud resulting from the final list of 291 different attributes.

Font sizes represent frequency, i.e. the bigger the font is, the more frequent was the attribute. Color of the attributes is used just for decoration and has no meaning.

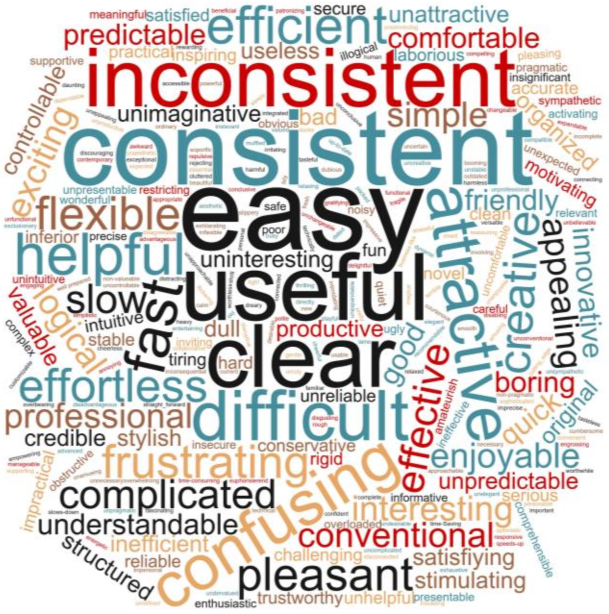

Figure 1: Word cloud of attributes used in the questionnaires.

The pure number of different attributes and the word cloud in Figure 1 show how divers and heterogenous the concept of user experience is.

If we look at the most prominent attributes (the words with biggest font size), we see that most of them are either pure valence items (e.g. easy, pleasant, attractive) or correspond to classical usability criteria, for example *Efficiency*, *Controllability* or *Learnability* (e.g. consistent, inconsistent, fast, useful, clear, difficult, confusing).

The high frequency of valence items results from the fact that they just describe a general impression towards a product and can thus be applied in many situations, i.e. they are used in many items.

The high frequency of usability related attributes results from the fact that a majority of the UX questionnaires from our list were created at a time where non-task related UX criteria were widely ignored in UX research and practice. Thus, these older questionnaires are dominated by items relating to pragmatic quality.

4 Item formats

If we look at the concrete items, we see that the most frequent item format are short statements for which the participant can express his or her level of agreement or disagreement on a Likert-scale.

Example from SUS [24]:

I found the system unnecessarily complex
Strongly disagree o o o o o *Strongly agree*

Example from ISONORM [11]:

The software	---	--	-	-/+	+	++	+++	The software
requires unnecessary inputs	O	O	O	O	O	O	O	does not require unnecessary inputs

This item format was used in 35 of the 40 questionnaires from the list, i.e. clearly dominates in current questionnaires.

An item format used in some questionnaires (4 from our list) is semantic differentials. Here an item consists of a pair of adjectives with opposite meanings. A participant can express his or her impression concerning the product on a semantic dimension.

Example from the UEQ [25]:

boring o o o o o o o *exciting*

The simplest item format is used in the MSPRC [13]. Here the items are just simple adjectives (for example *intuitive*, *complex*, *fast* or *valuable*) that can be associated with a product.

With two exceptions all questionnaires use an uneven number of answer categories. Here 21 have 7 answer categories and 13 use 5 categories. There is of course a tradeoff between the ability to express differences in the subjective impression and the cognitive complexity to decide between the offered answer categories. Most questionnaire authors seem to assume that 7 is the best tradeoff here.

A newer study [46] compared the impact of different numbers of response categories on the results of a questionnaire. The results showed that if the number of response categories reaches a certain minimal length (>5) this number has a limited impact on the results.

5 Measurement concept

There is a trivial dependency between the number of items in a scale and the accuracy of measurement. The more items we have, the lower is the impact of random response errors in single items on the scale mean. In addition, a single item can typically not express the meaning of complex UX concepts, like ease of learning, controllability or aesthetic impression. The more items we have in a scale, the better can the desired meaning of a scale be expressed by the combination of items.

On the other hand, the more items we have, the higher is the effort for the participants in a study to fill out the questionnaire and thus the completion time. Thus, we have here a classical tradeoff between required completion time and accuracy of measurement.

If we look at the questionnaires from our list, we can easily see that they follow different philosophies. Several questionnaires are obviously built with the goal to measure concepts as exact as possible by accepting long completion times. For example, ISOMETRICS used 75 items distributed to 7 scales to measure the usability of a product close to the definition in ISO 9241. PUTQ uses 100 items in 8 scales.

Other questionnaires are designed with the goal to capture just a rough impression of a participant towards a product with scales that require not much time to respond and can thus be applied in many practical settings. Examples are the SUS with 10 items used to calculate an overall usability score or the UMUX with just 4 items. Some questionnaires try to achieve a shorter response time for participants by using semantic differentials, for example UEQ with 4 items per scale or AttrakDiff2 with 7 items per scale.

Another clearly visible difference between items of the questionnaires from our list lies in the way the items are formulated.

Some questionnaires use quite concrete item formulations, for example:

- Is the ordering of menu options logical? (PUTQ)
- Messages always appear in the same place. (ISOMETRICS)
- The links provided in the material are clearly visible and logical. (WEBLEI)
- The software documentation is very informative. (SUMI)

This concrete item formulations make it easier to answer the items and is quite stable against misinterpretations. On the other hand, it restricts the applicability to a small group of products. For example, assume that a self-service application is evaluated. Such an application must be intuitive to use and will thus have no documentation. In such a case the SUMI question above will sound silly.

The opposite practice is to use items that just describe abstract impressions concerning a product. All items in semantic differentials are of this form, for example:

- useless / useful (UEQ+)
- lacking style / stylish (AttrakDiff2)

But this is not only possible with items from semantic differential:

- I felt very confident using the system. (SUS)
- The information is of high quality. (Web-Clic)

This way to formulate items requires a bit of abstraction from the participant, since the concrete interaction with the product must be mentally mapped on these more abstract impressions. It also increases the probability of misinterpretations based on the concrete research setting. For example, an item *insecure/secure* will be interpreted not exactly in the same way in an evaluation of a social network and an evaluation of a business software.

But on the other hand, such items are not directly tied to concrete elements in the user interface of a product and questionnaires using this approach can be applied in a wider range of products than questionnaires with very concrete items.

6 Scale names and semantic meaning

The scale names are often used as a first orientation concerning the UX aspects that a questionnaire tries to measure. However, they can be quite misleading. We illustrate this with a few examples.

AttrakDiff2 [1] and UEQ [25] contain both a scale named *Stimulation*. The corresponding items in the UEQ scale are: *boring / exiting*, *not interesting / interesting*, *motivating / demotivating*, *valuable / inferior*. The items in the AttrakDiff2 scale are: *original / conventional*, *unimaginative / creative*, *bold / cautious*, *innovative / conservative*, *dull / absorbing*, *harmless / challenging*, *novel / conventional*.

Both conceptualizations are similar, but of course not identical. The UEQ scale defines stimulation in the sense of an interesting and exiting interaction. This aspect is also contained in the corresponding AttrakDiff2 scale, but here in addition creativity and novelty of the design are seen as a part of *Stimulation*. In the UEQ this aspect is contained in a separate scale *Novelty*. Thus, semantically the combination of the UEQ scales *Stimulation* and *Novelty* corresponds to the AttrakDiff2 scale *Stimulation*.

Clearly both ways to operationalize the concept *Stimulation* are valid. A novel and creative design raises interest in a product and thus makes it more interesting. Thus, it is a valid approach to see *Novelty* as a part of *Stimulation*, but it is also a valid approach to keep both concepts separately, i.e. in two different scales.

But from the point of view of a UX professional that searches for a questionnaire this is of course an issue.

Another example is the concept of *Usefulness*. Corresponding scales are contained, for example, in meCUE [12] (scale name *Usefulness*), USE [31] (scale name *Usefulness*) and PUEU [16] (scale name *Perceived Usefulness*). In PUEU the concept is understood clearly as usefulness for the job, i.e. seen in a purely professional context. Example items from the corresponding PUEU scale are *Using the system would make it easier to do my job*, *Using the system would improve my job performance* or *Using the system in my job would increase my productivity*. In the corresponding scale of the USE the concept is understood much broader, as can be seen by items like *It gives me more control over the activities in my live* or *It does everything I would expect it to do*, but there are also statements close to the PUEU concept, for example *It helps me be more effective* or *It helps me to be more productive*. In the meCUE the items of the scale *Usefulness* are formulated more neutral, for example, *I consider the product extremely useful* or *With the help of this product I will achieve my goals*.

Thus, we see three different conceptualizations of *Usefulness* that are to some extend similar, but of course far away from being the same concept. Thus, it is not unlikely that we measure different values if we evaluate the same product with the same target group, but these three different scales.

Another example that shows an interesting aspect is the VISAWI [33]. This questionnaire concentrates on the measurement of visual aesthetics of websites (however, it also works quite well in other domains, for example to measure the visual appeal of user interfaces in business software). The questionnaire contains 4 scales that reflect different important aspects concerning aesthetic impression or beauty of a website. Some items, for example, *The design is uninteresting* or *The layout is pleasantly varied* reflect semantically aspects that relate closely to concepts like *Stimulation* or *Fun of Use* in other questionnaires. Items like *The layout is inventive* relate to concepts like *Novelty* and items like *The layout appears professionally designed*, or *The site is designed with care* are quite similar to scales like *Value* or *Identity*. It is important to notice that we do not find here a real one-to-one correspondence to other scales. In the scales of the VISAWI we find items semantically similar to existing scales in other questionnaires, but they are combined with new items related to visual appearance to form new scales.

The current situation with many scales that have unclear names and are to some extent semantically similar to scales in other questionnaires results from the common practice to create such scales empirically.

The process typically starts with a larger item pool that covers the UX aspects the questionnaire should measure. Then several products are evaluated by a larger group of participants with all those items. Over a statistical technique, for example, main component analysis, factors (hypothetical UX aspects corresponding to the scales) are extracted and the items that show the highest loadings on a factor are then selected to form the scale. The scale name is chosen by the researcher to somehow describe the common meaning of all the items in the scale.

This has the clear advantage that the resulting questionnaires cover with their scales the most relevant aspects of the domain under investigation. Items that are not relevant for the UX impression are simply sorted out by this approach.

But this way to construct UX questionnaires has also some inherent problems. The constructed scales will often consist of items that show high correlations but are from a purely semantical perspective not very homogeneous. In addition, the scale structure that is constructed depends on the products used in the data collection. Each item in a UX questionnaire is always interpreted in the context of the evaluated product. For example, an item *costs time / saves time* has a different meaning if a business software is evaluated (in this context it will show a high correlation to other items that represent efficiency) or if a social network is evaluated (in that context some participants will interpret it in the sense of the danger to spend too much valuable time in the network and the correlation to efficiency items may be much lower here). Thus, the correlation of items depends on the products used in the data collection (and for this reason it is a good idea to use different products here) and these correlations influence the selection of scales and items.

To sum up, the real meaning of a scale lies solely in the items. The scale name can be quite misleading and, in many cases, covers only parts of the meaning of a scale. This makes it quite difficult and cumbersome to find out if a questionnaire is really a good choice for a given research question. In addition, it is nearly impossible to relate findings obtained with different questionnaires.

7 How do questionnaires relate to each other?

In this section we try to describe which UX aspects are covered by the questionnaires in our list. As we have argued before, it does not make much sense to base such an analysis on the names and descriptions of the scales. Such an analysis must be done on the level of concrete items.

To be able to compare different UX questionnaires concerning the semantic aspects covered by their items, we need a common classification of relevant UX aspects. We use for the following analysis a set of 16 UX aspects [41, 42] that were extracted from an analysis of existing UX questionnaires and for which the relevance for different product categories was investigated in several studies. Each of these UX aspects represents a semantic UX quality and is described by a label and a short text.

The following UX aspects are contained: *Content Quality*, *Customization*, *Perspicuity*, *Efficiency*, *Immersion*, *Intuitive Usage*, *Usefulness*, *Novelty*, *Beauty*, *Identity*, *Controllability*, *Stimulation*, *Clarity*, *Loyalty*, *Trust*, and *Value*. Please check [42, 44] for a detailed description of their exact meaning.

Examples:

- *Efficiency:* I can achieve my goals with minimal time and minimal physical effort. The product responds quickly to my input.
- *Beauty:* The product is beautiful and attractive.
- *Usefulness:* Using the product brings me advantages. It saves me time and effort and makes me more productive.

This set of UX aspects is also used as a basis for an expert review method [43] or comparisons of different cultures [44] concerning the importance of UX aspects for certain product categories.

For the following analysis we counted for each of the questionnaires in our list how many items represent each of the 16 UX aspects. Some items fit equally well to two aspects. These are counted then with 0.5 in each of the aspects. All items that correspond to more than three UX aspects or to none of them were assigned to a category “Others”. Now we divide per questionnaire and UX aspect the number of items representing this aspect by the total number of items in the questionnaire. The resulting number shows how well the UX aspect is represented in a measurement by this questionnaire.

By calculating the Euclidian distance between two questionnaires we get a distance matrix of all 40 questionnaires.

We use multi-dimensional scaling (MDS) to visualize these data [45]. An MDS is based on a set of objects (in the case of this research the questionnaires) and a matrix that shows the similarity or distance for each pair of objects (in our case the distance matrix described above). The MDS then illustrates the objects as points in a two-dimensional space, so that the Euclidean distance between the points reflects the similarity of the objects as close as possible. Thus, it is mainly a visualization technique.

The MDS representation in Figure 2 allows some nice interpretations. In the middle (green font) we find with the UEQ+ (16 modular scales with 64 items) and MSPRD (118 attributes that are used to describe UX) two frameworks that are designed to cover a huge range of UX aspects. In addition, WEBQUAL (9 scales with 36 items) is placed here, which also shows a wide distribution of items over many UX aspects.

On the left side in the middle (dark blue font) we see a larger group of questionnaires that have a strong focus on the pragmatic quality aspects *Efficiency*, *Perspicuity*, *Intuitive Use*, and *Controllability* and contain in addition some pure valence items.

The questionnaires on the left side on top (light blue font) also have a strong focus on pragmatic aspects, but in addition also on *Usefulness*.

On the bottom of the left side (pink font) we see a group of questionnaires that also strongly emphasize on pragmatic quality, but have in addition many items representing *Content Quality* and *Clarity*. WEQ and WEBLEI also have their focus on pragmatic aspects and content quality, but do not take *Clarity* into account.

If we go to the right, we find questionnaires with a stronger focus on non-task related or hedonic UX aspects. To the left on bottom, there is the VISAWI, which concentrates purely on beauty (visual aesthetics). A little bit to the right there is the AttrakDiff2 that has 7 items concerning pragmatic quality, but 21 concerning *Attractiveness*, *Stimulation* and *Identity*.

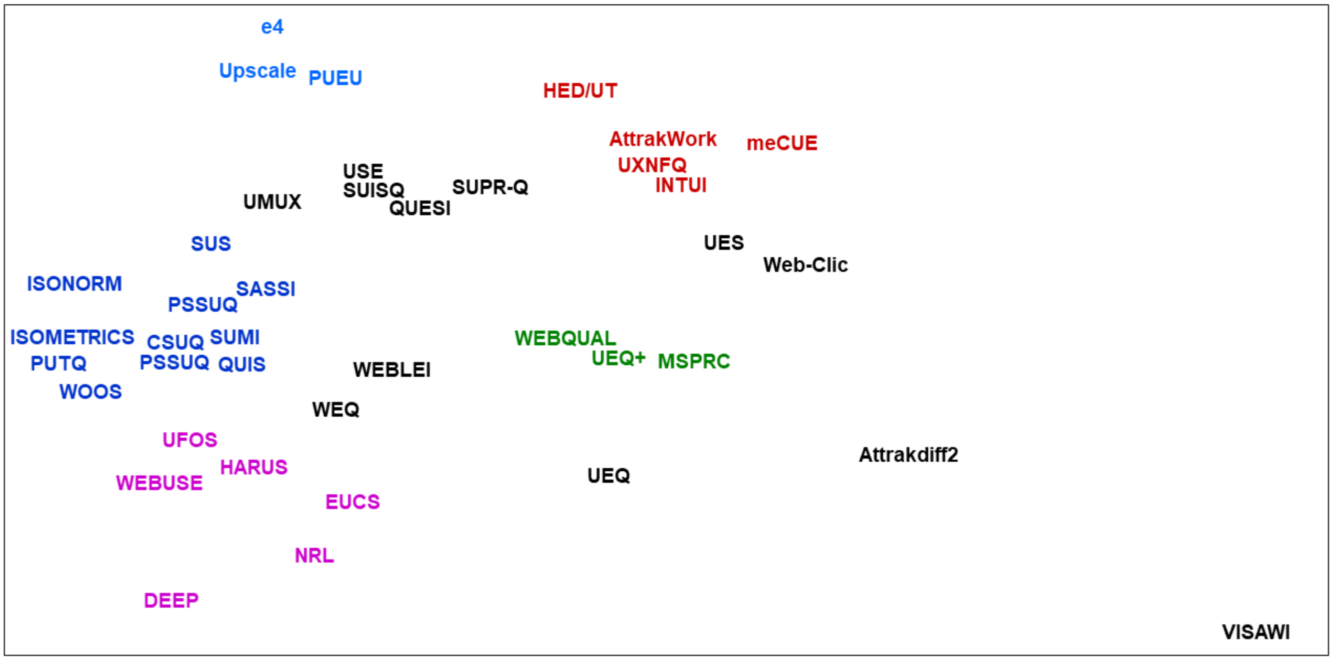

Figure 2: Multi-dimensional scaling of the questionnaires concering their semantics.

Above there is a group of questionnaires (red font) that have many items that measure non-task related qualities, especially *Stimulation*.

Thus, we can interpret the horizontal axis from left to right as a dimension representing a shift from the measurement of purely pragmatic UX aspects to purely hedonic UX aspects.

Of course, the assignment of the items to the scales is just done by the author and others may come to a different classification (for many items it seems clear, but there are of course items where there is room for discussion and different opinions may be possible). In addition, other schemes to group the items may be developed. Thus, other pictures concerning the semantic grouping of the UX questionnaires will be possible and are of course equally justified. But the current result is a first step to get a deeper understanding how existing UX questionnaires relate to each other and also show a method to derive such insights.

8 Summary

If we look at the current situation in the field of UX questionnaires from the point of view of an UX professional, we must state that it is highly confusing.

We have a huge number of existing questionnaires, each of them tailored to measure a specific subset of UX aspects. Some of them are generally applicable to many product categories and some of them are highly specialized for specific types of products (for example, websites) or product properties (defined by the scales). Due to different item formats and the different measurement philosophy it is difficult to combine these questionnaires in a single UX evaluation. First, it is of course time consuming for the participants to fill out many questions, second it is also confusing if they must adapt to several item formats in one investigation.

In addition, even if you get results with different questionnaires it is hard to find an overall interpretation of the results, for example due to a different scale format. In addition, different methods that help to interpret the results are used in the questionnaires (for example, different methods to set up a benchmark).

Modular questionnaires, like the UEQ+, try to solve this by providing several scales that can be combined by the researcher to form a questionnaire tailored to their research question. But a downside of this approach is that the modularity makes it hard to provide good supporting material, for example a benchmark, that helps to interpret the results.

What can be done to improve the current situation? Researchers that create UX questionnaires spend a lot of effort on empirical work to show that their scales are reliable and valid, thus naturally concentrate their work on getting a better understanding of their own questionnaires. What is missing to a large extent is empirical research that compares measurements of the same product with different questionnaires.

In addition, when new questionnaires are published authors should spend more effort to explain what their scales mean semantically. Just reporting a scale name and the items per scale is not enough to help UX professionals to decide if a questionnaire fits to their research questions.

REFERENCES

[1] Hassenzahl, M., Burmester, M., & Koller, F. (2004). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität In J. Ziegler & G. Szwillus (Hrsg.), Mensch & Computer 2003. Interaktion in Bewegung, pp. 187-196. Stuttgart, Leipzig: BG Teubner.

[2] Väätäjä, H., Koponen, T., & Roto, V. (2009). Developing practical tools for user experience evaluation: a case from mobile news journalism. In European Conference on Cognitive Ergonomics: Designing beyond the Product---Understanding Activity and User Experience in Ubiquitous Environments, pp. 23. VTT Technical Research Centre of Finland.

[3] Lewis, R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7(1), pp. 57-78.

[4] Yang, T., Linder, J., & Bolchini, D. (2012). DEEP: Design-Oriented Evaluation of Perceived Usability. International Journal of Human-Computer Interaction, 28(5), pp. 308–346.

[5] Harbich, S., Hassenzahl, M. & Kinzel, K. (2007). e4-Ein neuer Ansatz zur Messung der Qualität interaktiver Produkte für den Arbeitskontext. Oldenbourg Verlag.

[6] Doll, W. J., & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. MIS Quarterly, June, pp. 259–274.

[7] Santos, M. E. C., Polvi, J., Taketomi, T., Yamamoto, G., Sandor, C., & Kato, H. (2015). Toward standard usability questionnaires for handheld augmented reality. IEEE computer graphics and applications, 35(5), pp. 66-75.

[8] Van der Heijden, H., & Sørensen, L. S. (2003). Measuring attitudes towards mobile information services: An empirical validation of the HED/UT scale. In ECIS, pp. 765-777.

[9] Ullrich, D. (2014). Intuitive Interaktion: Eine Exploration von Komponenten, Einflussfaktoren und Gestaltungsansätzen aus der Perspektive des Nutzererlebens. Doctoral dissertation, Technische Universität Darmstadt.

[10] Gediga, G., Hamborg, K. C., & Düntsch, I. (1999). The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. Behaviour & Information Technology, 18(3), pp. 151-164.

[11] Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In Software-Ergonomie'97, pp. 253-262. Vieweg Teubner Verlag.

[12] Minge, M. & Riedel, L. (2013). meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. Mensch & Computer 2013: Interaktive Vielfalt, pp. 89-98.

[13] Benedek, J., & Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. Proceedings of Usability Professionals Association, 2003(8-12), 57.

[14] Thielsch, M. T., Blotenberg, I. & Jaron, R. (2014). User evaluation of websites: From first impression to recommendation. Interacting with Computers, 26 (1), 89-102.

[15] Lewis, J. R. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 36, No. 16, pp. 1259-1260. Sage CA: Los Angeles.

[16] F.D. Davis (1989). Perceived Usefulness, Percevied Ease of Use, and User Acceptance of Information Technology, MIS Quarterly, Vol.13, pp.319-340.

[17] Lin, H. X., Choong, Y. Y., & Salvendy, G. (1997). A proposed index of usability: A method for comparing the relative usability of different software systems. Behaviour & Information Technology, 16(4-5), pp. 267-277.

[18] Hurtienne, J. & Naumann, A. (2010). QUESI—A questionnaire for measuring the subjective consequences of intuitive use. Interdisciplinary College, pp. 536.

[19] Chin, J.P., Diehl, V.A. & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human–computer interface. In: Proceedings of CHI 1988, ACM, Washington, DC, pp. 213-218.

[20] Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). Natural Language Engineering, 6(3-4), pp. 287–303.

[21] Polkosky, M. D. (2008). Machines as mediators: The challenge of technology for interpersonal communication theory and research. In: E. Konjin (Ed.), Mediated Interpersonal Communication, pp. 34–57. New York, NY: Routledge.

[22] Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. British Journal of Educational Technology, 24(3), pp. 210-212

[23] Sauro, J. (2015). SUPR-Q: A comprehensive measure of the quality of the website user experience. Journal of Usability Studies, 10(2), pp. 68-86.

[24] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), pp. 4-7.

[25] Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and Usability Engineering Group, pp. 63-76. Springer, Berlin, Heidelberg.

[26] Schrepp, M., & Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. International Journal of Interactive Multimedia & Artificial Intelligence, 5(7).

[27] O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies, 112, pp. 28-39.

[28] Konradt, U., Wandke, H., Balazs, B., & Christophersen, T. (2003). Usability in online shops: scale construction, validation and the influence on the buyers' intention and decision. Behaviour & Information Technology, 22(3), pp. 165-174.

[29] Finstad, K. (2010). The usability metric for user experience. Interacting with Computers, 22(5), pp. 323-327.

[30] Karlin, B. & Ford, R. (2013). The Usability Perception Scale (UPscale): A measure for evaluating feedback displays. In: Proceedings of the 2013 Human Computer Interaction (HCII) Conference. Las Vegas, NV: ACM.

[31] Lund, A. (2001). Measuring usability with the USE questionnaire. In: Usability and User Experience Newsletter, STC Usability SIG, 8(2), pp.1–4.

[32] Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. Journal of Personality and Social Psychology, 89, pp. 325–339.

[33] Moshagen, M. and Thielsch, M.T. (2010). Facets of visual aesthetics. International Journal of Human-Computer Studies, 68, pp. 689–709.

[34] Kirakowski, J. & Cierlik, B. (1998). Measuring the usability of web sites. In Proceedings of the Human Factors and Ergonomics Society annual meeting, 42(4), pp. 424-428. Sage CA: Los Angeles.

[35] M. Thielsch & G. Hischfeld (2019): Facets of Web-Site Content. Human-Computer Interaction, 34(4), pp. 279-327.

[36] Chang, V. (1999). Evaluating the effectiveness of online learning using a new web based learning instrument. Proceedings Western Australian Institute for Educational Research Forum.

[37] Barnes, S. & Vidgen, R. (2000). WebQual: An exploration of website quality. ECIS 2000 Proceedings, 74.

[38] Chiew, T. K., & Salim, S. S. (2003). Webuse: Website usability evaluation tool. Malaysian Journal of Computer Science, 16(1), pp. 47-57.

[39] Elling, S., Lentz, L. & De Jong,M. (2007). Website Evaluation Questionnaire: Development of a research-based tool for evaluating informational websites. Lecture Notes in Computer Science, 4656, pp. 293–304.

[40] Yom, M., Wilhelm, T. (2004). WOOS – Ein Messinstrument für die wahrgenommene Orientierung in Online-Shops. In: R. Keil-Slawik, H. Selke, G. Szwillus (Hrsg.): Mensch & Computer 2004: Allgegenwärtige Interaktion. München: Oldenbourg Verlag, pp. 43–53

[41] Winter, D., Schrepp, M. & Thomaschewski, J., (2015). Faktoren der User Experience - Systematische Übersicht über produktrelevante UX-Qualitätsaspekte. In: Endmann, A., Fischer, H. & Krökel, M. (Hrsg.), Mensch und Computer 2015 – Usability Professionals. Berlin: De Gruyter Oldenbourg, pp. 33-41.

[42] Winter, D., Hinderks, A., Schrepp, M. & Thomaschewski, J. (2017). Welche UX Faktoren sind für mein Produkt wichtig? In: Hess, S. & Fischer, H. (Ed.), Mensch und Computer 2017 - Usability Professionals. Regensburg: Gesellschaft für Informatik e.V., pp. 191 – 200.

[43] Held, T., Schrepp, M. & Mayalidag, R. (2019). User Experience Review - Ein einfaches und flexibles Verfahren zur Beurteilung der User Experience durch Experten. In: Fischer, H. & Hess, S. (Hrsg.), Mensch und Computer 2019 - Usability Professionals. Bonn: Gesellschaft für Informatik e.V. und German UPA e.V.

[44] Santoso, H.B. & Schrepp, M. (2019). The Impact of Culture and Product on the Subjective Importance of User Experience Aspects. Helion, 5(9).

[45] Torgerson, W.S. (1958). Theory and Methods of Scaling. Wiley, New York.

[46] Lewis, J.R. & Erdinc, O. (2017). User Experience Rating Scales with 7, 11, or 101 Points: Does it matter? Journal of Usability Studies, 12(2), pp. 73-91.