

Developing a UX KPI based on the User Experience Questionnaire

Andreas Hinderks^{a,*}, Martin Schrepp^b, Francisco José Domínguez Mayo^a, María José Escalona^a, Jörg Thomaschewski^c

^aUniversity of Seville, Spain

^bSAP AG, Germany

^cUniversity of Applied Science Emden/Leer

Abstract

Decisions in Companies are made typically by using a number of entirely different key figures. A user experience key figure is one of many important key figures that represents one aspect of the success of the company or its products. What we aim in this article is to present to those responsible for a product a method of how a user experience key performance indicator (UX KPI) can be developed using a UX questionnaire. We have developed a UX KPI for use in organizations based on the User Experience Questionnaire (UEQ). To achieve this, we added six questions to the UEQ to measure the importance of the UEQ scales. Based on the UEQ scales and the scores given for importance, we then developed a User Experience Questionnaire KPI (UEQ KPI). In a first study with 882 participants, we calculated and discussed the UEQ KPI using Amazon and Skype. The results show that the six supplementary questions could be answered independently of the UEQ itself. In our opinion, the extension can be implemented without any problems. The resulting UEQ KPI can be used for communication within an organization as a key performance indicator.

Keywords: User experience, Usability, KPI, Management, User Experience Questionnaire (UEQ)

1. Introduction

In many companies, decisions are mostly made based on key figures, such as turnover, profit, or employee satisfaction. Managers are accustomed to information being summarized and available via key figures.

It's often difficult for a business to use just one key figure for the entire product. Decision makers in companies use typically a number of quite different key figures. Each key figure represents one aspect that is important for the success of the company or its products. For this reason, each department that contributes to the success of a product typically has an individual key figures that reflects the contribution of the department. Monitoring can be carried out based on these key figures to better direct the company [1]. This leads to the demand to map measurement results concerning UX into a single number, i.e. a key performance indicator (KPI).

User experience is generally understood as a multidimensional construct [2]. For example, in order to obtain a good user experience, a product should be easy to learn, efficient to use or well controlled with additional criteria like aesthetics, joy-of-use, novelty or attractiveness. The first group of criteria is often referred as pragmatic quality

aspects [3], while the second group is called hedonic quality aspects. Another often-used terminology to distinguish both classes of quality criteria is usability goals versus user experience goals [4].

A well-known definition of user experience is given in ISO 9241-210 [5]. Here user experience is defined as “a person's perceptions and responses that result from the use or anticipated use of a product, system or service”. Thus, user experience is seen as a holistic concept that includes all types of emotional, cognitive or physical reactions concerning the concrete or even only the assumed usage of a product formed before, during and after use. However, the standard does not provide a definite list of factors or methods to measure user experience. A different interpretation is to define user experience as a set of distinct quality criteria [4] that includes classical usability criteria, like efficiency, controllability or learnability, and non-goal directed or hedonic quality criteria [3], like stimulation, fun-of-use, novelty, emotions [6], or aesthetics [7]. This has the advantage that it splits the general notion of user experience into a number of simple quality criteria, which describe distinct and relatively well-defined aspects of user experience that can be measured independently.

Questionnaires that measure the user experience take into account this complexity of user experience, since they usually compute values on different UX scales. A scale corresponds to a content-delimited quality characteristic of user experience, e.g., efficiency or originality. Depending on the questionnaire, other combinations of quality characteristics are measured. Standardized questionnaires are

*Corresponding author

Email addresses: andreas.hinderks@iwt2.org (Andreas Hinderks), martin.schrepp@sap.com (Martin Schrepp), fjdominguez@us.es (Francisco José Domínguez Mayo), mjescalona@us.es (María José Escalona), joerg.thomaschewski@hs-emden-leer.de (Jörg Thomaschewski)

not a more or less random or subjective collection of questions, but result from a careful construction process. This process guarantees accurate measuring of the intended UX qualities. But on the other hand, a standard UX questionnaire is not able to measure user experience holistic [8]. A standardized questionnaires accurately measures the UX scales identified in the construction like stimulation, efficiency, attractiveness, etc.

A UX designer is usually interested in the strengths and weaknesses of the product. A questionnaire, which results in a series of scale values, meets this need for detailed information. For managers or decision makers, this is usually too complex. First, the importance of each UX quality aspect for the product is often poorly understood. Second, managers and decision-makers are more interested in the question of whether there have been improvements or deteriorations compared to the last UX measurement. This is easier to decide if you only need to compare a single KPI and not a number of individual UX scales. Hence, there is also a strong desire for a single user experience key figure.

The method presented in this paper is based on the User Experience Questionnaire (UEQ) [9] and shows how to get a KPI by extending the questionnaire. We decided to extend the UEQ because it is a well-known UX questionnaire, it is available in Spanish, English, and German, and can measure pragmatic and hedonic qualities. The original version of the UEQ was designed in German and English [9], but has so far been translated to several languages like Spanish [10] and Portuguese [11]. The objective of the UEQ is to allow a quick assessment done by end users covering a preferably comprehensive impression of user experience. It should allow the users to express feelings, impressions, and attitudes that arise when experiencing the product under investigation in a very simple and immediate way. It consists of 26 items that are grouped into 6 scales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty). Each scale represents a distinct UX quality aspect.

In this article, we will develop a method for a user experience KPI by six additional questions of the User Experience Questionnaire (UEQ). The result is a UX KPI. In a first Evaluation with Amazon and Skype, we will interpret and discuss the UX KPI.

Section 2 surveys the works related to key performance indicators and user experience questionnaires. Section 3 outlines our method to develop a user experience KPI. Furthermore, we describe a first study to evaluate our method. In Section 4 we present the results of our study. Section 5 discusses the results of our study.

2. Background and related work

There are many different metrics and KPIs. In one study, Alves et al. [12] found that about 18% of the respondents already used a UX or usability KPI. The study also shows that other methods and tools in fields of knowledge involved in UX evaluations such as questionnaires (34%),

cognitive walk-throughs (40%), interviews (58%), and observation (63%) are more commonly used than a KPI.

A KPI is a metric, but a metric is not necessarily a KPI [13]. This means that there are many metrics that claim to deliver KPIs. Metrics are tools that help facilitate decision-making processes and improve organization performance [14]. Basically, each metric is therefore suitable as a KPI. But it is crucial that the organization be familiar with and use the metric.

As noted in the introduction, UX questionnaires are a common quantitative measure of user experience [15]. User experience questionnaires are characterized by the fact that they measure the subjective attitude of the user towards the test object. The respondent evaluates statements (items), for example by selecting a suitable category from the value range of a rating scale.

A number of UX questionnaires exist. Each questionnaire contains different scales for measuring groups of UX aspects. Questionnaires that measure pure usability aspects are, for example, the System Usability Scale (SUS) [16] and the Software Usability Measurement Inventory (SUMI) [17].

Questionnaires covering the broader aspect of UX are, for example, the Visual Aesthetics of Websites Inventory (VisAWI) [18], the User Experience Questionnaire (UEQ) [9], and Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) [19].

In principle, UX questionnaires with a total value can be divided into two different categories: questionnaires with only one overall result and questionnaires that calculate the overall result as an average of sub-scales. The list below are examples of UX questionnaires that only provide a total score. Some of the questionnaires contain only a single item, so the response made to that item equals the overall result.

- **Brooke 1986 [16]:** System Usability Scale (SUS) with ten items
- **Lewis 1995 [20]:** After Scenario Questionnaire (ASQ) with three items
- **Reichheld 2003 [21]:** Net Promoter Score (NPS) with one item
- **Sauro et al. 2009 [22]:** Single Ease Question (SEQ) with one item
- **Sauro et al. 2009 [22]:** Subjective Mental Effort Questionnaire (SMEQ) with one item

On the other hand, the following two questionnaires are designed to provide results for subsidiary scales as well as overall results. This overall result is calculated by finding the arithmetic mean of the sub-scales.

- **Lewis 2002 [23]:** Post-Study System Usability Questionnaire (PSSUQ) with 19 items und three sub-scales

- **Moshagen et al. 2010 [18]:** Visual Aesthetics of Websites Inventory (VisAWI) with 18 items and four sub-scales

The overall result as well as the sub-scale results are generally usable as KPIs, since these are valid as metrics.

Another method has been used by Sauro et al. [24], using four different scores (task completion, error counts, task times, and satisfaction scores) to create a uniform and standardized usability metric. This metric was calculated using both the individual scores and the weighting of the coefficients from a principal components analysis. The amount of information provided in the resulting metric is higher than the mean value from the individual sources.

A well-used UX KPI should take into account different meanings of different criteria in relation to different scenarios and user groups to provide an interpretable value. The VisAWI and PSSUQ calculate a KPI over the average of the individual sub-scales. No importance of individual quality characteristics is taken into account in the calculation.

The importance of individual quality characteristics for the overall impression will generally vary between different products and user groups, as well. For example, an efficient operation would be much more important for a product that is used multiple times a day than for a web service that is used only once a year. Conversely, an intuitive operation would be central to this web service, but not for the application used daily.

A simple possibility of a KPI based on ux questionnaire would be to simply calculate the mean of all items or scale and use that value as a KPI. The basic question, however, is what a KPI generated in this way actually means. We therefore consider this approach to be an inadequate method. In the next chapter we describe a method that also takes into account the individual quality characteristics.

3. Research methodology

In this section, we will describe our approach in detail. Our approach is divided into three different steps:

1. Step 1: Additional six questions to the User Experience Questionnaire to measure the perceived importance of each UEQ scale (see 3.1).
2. Step 2: Method to develop a UX KPI using the variables from step 1 (see 3.2).
3. Step 3: First Evaluation of the method from step 2 by conducting a study with Amazon and Skype (see 3.3).

The different steps are explained in more detail in the next three paragraphs.

3.1. Additional questions to the UEQ

The central idea of our approach is to additionally measure, per UEQ scale, how important the participant is to the scale in relation to the product to be assessed. The additional questions establish a direct connection to the scales by mentioning the most important UEQ items (each in the positive form) of the corresponding scale. Because of this connection, we can assume that the question really captures the importance of the associated scale.

For this purpose, additional questions were developed for the six UEQ scales. The list of English questions are as follows:

- **Attractiveness:** The product should look attractive, enjoyable, friendly and pleasant.
- **Efficiency:** I should perform my tasks with the product fast, efficient and in a pragmatic way.
- **Perspicuity:** The product should be easy to understand, clear, simple, and easy to learn.
- **Dependability:** The interaction with the product should be predictable, secure and meets my expectations.
- **Stimulation:** Using the product should be interesting, exiting and motivating.
- **Novelty:** The product should be innovative, inventive and creatively designed.

The English, Spanish and German versions of the questions are listed in Table 1.

3.2. Developing the UEQ KPI

The formula described here is to combine for every participant the value and assessed importance of each UEQ scale. Therefore, we calculated for each participant the relative importance of each scale and multiplied this with the value of the UEQ scale. This allows to generate a meaningful UX KPI based on the UEQ, that we have named UEQ KPI. The steps are described in detail below.

- The result of the UEQ represents a value per participant Pa_i and for every scale: attractiveness A_i , perspicuity P_i , efficiency E_i , dependability D_i , stimulation S_i , and novelty N_i .
- The result of the six additional questions is a value of the assessed importance of each scale: attractiveness a_i , perspicuity p_i , efficiency e_i , dependability d_i , stimulation s_i , and novelty n_i .
- We have calculated the relative importance per participant using the given assessed importance of each scale. For example for attractiveness $Aw_i = a_i / (a_i + p_i + e_i + d_i + s_i + n_i)$.

Table 1: Additional questions to assess the importance of the parameters to the participant.

Scale	Statements
Attractiveness	[ENG] The product should look attractive, enjoyable, friendly and pleasant.
	[SPA] El sistema tiene que ser atractivo, agradable y simpático.
	[GER] Das Produkt soll attraktiv, angenehm und sympathisch wirken.
Efficiency	[ENG] I should perform my tasks with the product fast, efficient and in a pragmatic way.
	[SPA] El sistema debe ayudar a que yo haga mis tareas de forma rápida, eficiente y pragmática.
	[GER] Das Produkt soll mir helfen, meine Aufgaben schnell, effizient und pragmatisch zu erledigen.
Perspicuity	[ENG] The product should be easy to understand, clear, simple, and easy to learn.
	[SPA] El sistema debe ser entendible, claro y fácil de amigable.
	[GER] Das Produkt soll übersichtlich, verständlich und leicht zu lernen sein.
Dependability	[ENG] The interaction with the product should be predictable, secure and meets my expectations.
	[SPA] El uso del sistema debe ser predecible de usar y fácil de controlar.
	[GER] Die Bedienung des Produkts soll in der Bedienung vorhersehbar und gut kontrollierbar sein.
Stimulation	[ENG] Using the product should be interesting, exiting and motivating.
	[SPA] Trabajar con el sistema debe ser interesante, emocionante y de motivador.
	[GER] Das Arbeiten mit dem Produkt soll interessant, spannend und aktivierend sein.
Novelty	[ENG] The product should be innovative, inventive and creatively designed.
	[SPA] El sistema debe ser original, innovador y creativo.
	[GER] Das Produkt soll originell, innovativ und kreativ gestaltet sein.

- The UEQ KPI per participant can be calculated by multiplying the mean value of the UEQ scale with its relative importance: $UEQ\ KPI_i = Aw_i * A_i + Pw_i * P_i + Ew_i * E_i + Dw_i * D_i + Sw_i * S_i + Nw_i * N_i$.

The UEQ KPI has been calculated using the following formula:

$$UEQ\ KPI = \frac{1}{n} \sum_{i=1}^n (Aw_i * A_i + Pw_i * P_i + Ew_i * E_i + Dw_i * D_i + Sw_i * S_i + Nw_i * N_i) \quad (1)$$

We evaluated this method to calculate the UEQ KPI in two studies with Amazon and Skype.

3.3. First evaluation of the UEQ KPI

The study is motivated by a need to understand the relationship between the assessed quality (i.e. the value measured on the corresponding UEQ scale) and the assessed importance. It can be assumed that these two assessments are not completely independent of each other. If a user considers a quality aspect to be irrelevant for a product, he will also assess the items on the corresponding scale less seriously. Because of this, we expect a certain dependence.

3.3.1. Object of study

In this study, products with a high level of awareness were evaluated to ensure the subjects could assess the products. The test objects selected were Amazon and Skype. We chose Amazon and Skype because we have

several datasets with these two products and different language versions of the UEQ published in various papers [10, 11, 25, 26]. We expect that there will be no significant differences between Amazon and Skype, the language version from this study, and the studies we have already published.

3.3.2. Purpose

The purpose of the study is to assess the individual assessment with the UEQ and the additional six questions about the importance of the UEQ scale. The study provides information on whether the participant performs the importance independently of the evaluation.

3.3.3. Quality focus

The main focus of the study is to evaluate the UEQ KPI by evaluating Amazon and Skype. Here, two specific aspects are emphasized. The choice is to focus on the scale consistency and the correlation between the value and importance of the UEQ scale.

3.3.4. Context

The study has been conducted in England, Spain, and Germany through online and paper versions of the questionnaire. For the English and German dataset, we have chosen a social panel (Prolific Academic¹) to collect the data. We have collected the Spanish dataset from the University of Seville.

First of all, the participants were asked to choose between Amazon and Skype. In the next step, they had to complete the UEQ [9]. Based on the six UEQ scales,

¹<https://www.prolific.ac>

the participants were asked in the following step to assess how important the scale and/or product quality is to them for the chosen product (completely unimportant with the poles/very important on a 7-point Likert scale). The additional questions were given after the evaluation of the product with the UEQ, so that the actual evaluation is not affected. Finally, the participants had to state their age and gender. Each participant only evaluated one product.

A total of 882 participants took part in the study. In addition to the UEQ, we also asked for age and gender. The participants assured us that they had used the chosen product at least once a month. Not all participants answered all items seriously. To detect the somewhat random or non serious answers and clean up the datasets, we used a simple heuristic [27]. In the end, we identified 87 answers as problematic and deleted them from the datasets.

The remaining answers are divided into 433 (Table 2) for Amazon and 362 for Skype (table 3).

Table 2: Number of Participants Choosing Amazon

Amazon		
England	238	(158 females, 96 males)
Spain	51	(no data available)
Germany	144	(52 females, 102 males)
Total	433	

Table 3: Number of Participants Choosing Skype

Skype		
England	215	(172 females, 82 males)
Spain	62	(no data available)
Germany	85	(40 females, 52 males)
Total	362	

The average age is 34 years (35 for females, 33 for males) for the English dataset and 29 years (31 for females, 28 for males) for the German dataset. We have no socio-demographic information available for the Spanish dataset.

4. Results

Thus, overall the participants had a slightly positive or neutral impression concerning the user experience of Amazon (Fig. 1) and Skype (Fig. 2). A detailed report of the study was created [28].

The impression concerning the pragmatic quality (Perspicuity, Efficiency and Dependability) is clearly higher than the impression concerning the hedonic quality (Stimulation, Novelty). During the evaluation, we did not find any significant differences between men and women.

Figure 1: Amazon - UEQ Scales (blue) and Assessed Importance (red) with 5% Confidence Interval as Error Bar

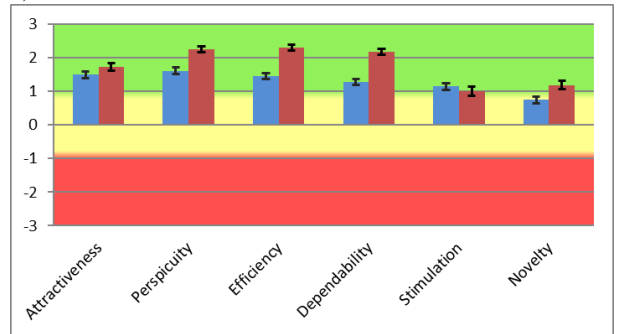
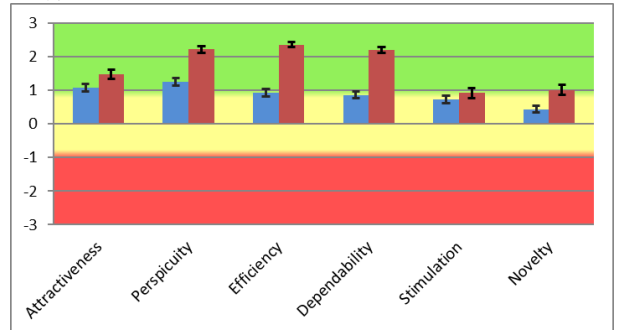


Figure 2: Skype - UEQ Scales (blue) and Assessed Importance (red) with 5% Confidence Interval as Error Bar



In Table 4, the values for each scale, country, and product are for the UEQ, respectively as well as the estimated importance.

The results from this study, without the results from the six additional questions, are mostly consistent with the results from earlier studies. It was found that participants rated Amazon and Skype differently for the same language version [10, 11]. Two further studies showed that there are differences between different language versions but that the same test object remains (Amazon and Skype) [25, 26]. In this respect, the results of the study are in line with expectations.

Reliability is typically estimated using the Standardized Cronbach Alpha coefficient [29]. The Cronbach's Alpha is a measure of the internal consistency of a questionnaire dimension [30]. An analysis of the Cronbach Alpha coefficient showed that the single scales showed high consistency values (Amazon: ATT: 0.90, PER: 0.81, EFF: 0.75, DEP: 0.67, STI: 0.78, NOV: 0.74 - Skype: ATT: 0.92, PER: 0.83, EFF: 0.80, DEP: 0.71, STI: 0.79, NOV: 0.74). This is an indicator that the scales are sufficiently consistent [29]. There is no general rule about how large the value should be. In practice, however, a value of > 0.7 has proven to be sufficient solidly established [31].

Overall, the assessed importance of pragmatic quality (Efficiency, Perspicuity, and Dependability) is higher than the value of the corresponding UEQ scale. In the case of the hedonic quality scales (Stimulation and Novelty) and

Table 4: UEQ Scales (UEQ) and Assessed Importance (IMP) for every Scale, Country and Product

	Amazon						Skype					
	ENG		SPA		GER		ENG		SPA		GER	
	UEQ	IMP	UEQ	IMP	UEQ	IMP	UEQ	IMP	UEQ	IMP	UEQ	IMP
Attractiveness	1.535	2.643	1.294	3.027	1.516	2.713	1.119	2.457	1.182	2.723	0.933	2.567
Perspicuity	1.710	3.149	1.069	3.308	1.649	3.363	1.331	3.263	1.165	3.156	1.126	3.186
Efficiency	1.641	3.439	1.074	3.208	1.276	3.102	1.045	3.539	0.683	3.254	0.794	3.072
Dependability	1.277	3.224	1.054	3.113	1.351	3.076	0.914	3.341	0.891	3.109	0.703	2.979
Stimulation	1.201	2.086	1.181	2.396	1.038	1.701	0.691	1.884	0.956	2.250	0.653	1.979
Novelty	0.825	2.322	0.951	2.566	0.542	1.841	0.407	2.129	0.560	2.328	0.453	1.866

the valence scale Attractiveness, the values are approximate. This was also expected of us, as both products are tools that are usually used frequently.

The total time to fill out the questionnaire was increased by about 50% compared to the UEQ. On average, it was 96 seconds for the UEQ (Amazon: 94 seconds and Skype: 98 seconds) and 42 seconds for the additional questions (Amazon: 39 seconds and Skype: 45 seconds).

The formula presented in Chapter 3 was used to calculate the UEQ KPI. Figures 3 and 4 show the UEQ KPI for England, Spain and Germany and the resulting total KPI for Amazon and Skype, respectively.

Figure 3: Amazon - KPI per Country

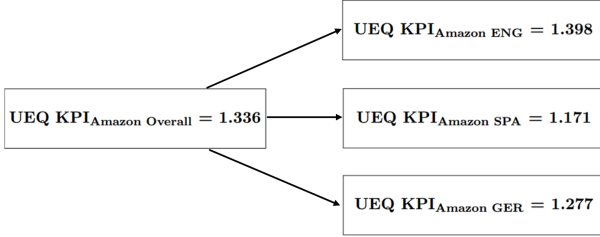
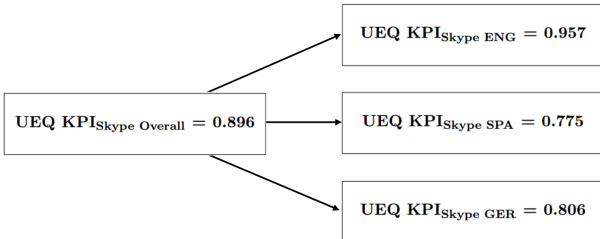


Figure 4: Skype - KPI per Country



The results show that the UEQ KPI are different per country and deviate from the overall UEQ KPI. Deviations with Amazon are +4.64% for England, -12.35% for Spain and -4.42% for Germany. In the case of Skype, the deviations for England are +6.81%, for Spain -13.50% and for Germany -10.04%.

The correlation between the UEQ scales and the importance per scale rating differs only slightly between Amazon and Skype (Table 5 and Table 6). Differences can be noted between the individual countries. In particular, the Spanish dataset differs from the other datasets.

Table 5: Amazon - Pearson correlation between Scale Means and Assessed Importance

Scale	Amazon - Pearson correlation			
	Overall	ENG	SPA	GER
Attractiveness	0.323	0.371	0.080	0.318
Perspicuity	0.206	0.263	0.181	0.113
Efficiency	0.247	0.368	-0.051	0.079
Dependability	0.217	0.297	-0.240	0.255
Stimulation	0.327	0.405	0.026	0.282
Novelty	0.320	0.299	0.115	0.367

Table 6: Skype - Pearson correlation between Scale Means and Assessed Importance

Scale	Skype - Pearson correlation			
	Overall	ENG	SPA	GER
Attractiveness	0,149	0,164	-0,032	0,211
Perspicuity	0,172	0,138	0,025	0,322
Efficiency	0,139	0,178	-0,071	0,107
Dependability	0,155	0,152	0,147	0,128
Stimulation	0,237	0,331	-0,137	0,155
Novelty	0,159	0,219	-0,034	0,139

In the case of Amazon, the correlation can be classified according to Cohen [32] as a mean ($r = .3$) and Skype as low ($r = .1$) correlation.

5. Discussion

According to the benchmark of the UEQ [33], the results of the UEQ are in the typical middle range compared. Overall, the hedonic qualities are valued lower than the pragmatic qualities. This is due to the fact that Amazon and Skype are launched products. The hedonic qualities change over time and are then rated more in a more neutral manner [34].

The participants were also able to submit a differentiated assessment of the importance of the UEQ scales. The results show that the evaluation of importance provides a different conclusion than the assessment of quality. The importance of pragmatic quality is much higher than the

importance of hedonic quality. That is understandable because Amazon and Skype are not necessarily experience-oriented but rather tools.

5.1. Correlation

The main focus of the study was to determine if the subjects could independently assess the six additional questions on the assessed importance of the UEQ scale. The correlations between the assessment of the importance of the UEQ scales show a medium to low correlation. This is in line with our expectations and means that the participants have completed this independently. The additional questions about importance have, therefore, had no major impact on the original UEQ.

From the low correlation it can be concluded that two independent judgments were formed: On the one hand, quality (UEQ) and on the other hand, importance (additional six items).

5.2. Cultural differences

The results from our two studies show different values for the UEQ scales and the importance of the scales for each country for the same product. These differences can be explained in part by different cultural factors [35]. Parsons et al. [36] have identified common social factors. Schwartz [37] found factors in the value structure worldwide. Despite identified universal factors across all cultures, Schwartz et al. [38] found that even these factors can differ in cultures.

Different cultures have different quality requirements for a product. In Germany, for example, reliability and functionality are expected first and foremost [39]. In the UK, a good image and robustness, as well as the price are important [39].

The focus of the study was not to determine the cultural differences, but to determine the robustness of the applied method, which we describe in the next section.

5.3. Possible value range of the UEQ KPI

The practical application of the UEQ KPI raises the question of whether a calculated value for the UEQ KPI is good or bad. In other words, to what extent does a UEQ KPI value represent a well or poorly measured user experience?

The formula for calculating the UEQ KPI consists of two parts for each factor of the UEQ: The first is the value of the UEQ factor and the second is the perceived importance per UEQ factor. The UEQ KPI is calculated from both values. In order to determine the value range of the UEQ KPI, both values should be considered.

A way to calculate the possible value range is to include the UEQ benchmark. A benchmark already exists for the UEQ, which specifies a practical value range for the factors of the UEQ [33]. The mean values of the benchmark categories can be used to calculate the possible range of values

of the UEQ KPI with practical data. All possible combinations of all benchmark categories and for each perceived importance factor (1 to 7 per UEQ factor) were calculated [40]. The result is a possible range for the UEQ KPI:

UEQ KPI value range between **-0.286** and **2.143**

Although a theoretical value range from -3 to +3 is possible, it will not occur in a practical application. Rather, it shows that the real value range is smaller and more positive.

5.4. Enhancement of the UEQ

The results from Section 4 suggest that the UEQ can be extended to gain the importance of the scales. In principle, this procedure should also work for other questionnaires, which contain several scales clearly separated from each other in content. The method described here does not lose any information. Although the UEQ has been extended with 6 additional questions, there is still a fully completed UEQ. This means that the information of the individual scales are still present and can be evaluated.

There is currently a larger benchmark for the UEQ scales [33], which helps to better interpret measured results. Such a benchmark has to be developed for the UEQ KPI. This will take more time since a larger amount of evaluation results has to be collected for a benchmark.

5.5. Limitations

In this article, we have focused on the design and evaluation of the UEQ KPI. The introduction of the UEQ KPI in an organization must be considered separately. The study carried out by Winter et al. [41] showed that there were several issues with both the construction of a usability KPI and the introduction of the KPI. The biggest problem was a lack of communication. Because of this, special attention should be given to the introduction of the UEQ KPI for a specific product as well as to the introduction. The UEQ KPI should be explained and discussed. It requires a holistic view of the organization's structures and processes so that the UEQ KPI is used as it was intended by the Management.

6. Conclusions and future work

We have presented a method for calculating user experience key performance indicator using the new UEQ KPI questionnaire. On the basis of the User Experience Questionnaire (UEQ), we queried respondents' perspectives on the importance of the UEQ scales with six additional items. The resulting UX KPI summarizes both the subjective rating of a product and the perceived importance of the UEQ scales in a KPI. We were able to evaluate this method in an initial study in three countries with overall 882 participants by evaluating Amazon and Skype.

In the literature, there are approaches to validate a KPI [42, 43]. In the end, the UEQ KPI has to be proved with different established and new products and language versions. The quality criteria for validation should be a low correlation between the scales of the UEQ and the six additional questions. Furthermore, the validation should show a stable result for the same product and language version but different studies. It should be noted that the results of the UEQ may vary depending on the product and language version [10, 11, 25, 26].

The extent to which the UEQ KPI can be used in practice depends on the product and the organization. The extent to which the UEQ KPI can be used for the desired application should be checked beforehand. Whether the UEQ KPI can serve as a metric must be validated in future applications with different products.

It is important to note that no information is lost through the method described. The UEQ has been expanded by six additional questions, but an entirely completed questionnaire is still available. This means that the information on the individual scales is still available. The questionnaire has only been extended and not changed in itself. Whoever is responsible for a product has the choice. On the one hand, they can use all the more detailed information that the UEQ offers. On the other side, they can use the UEQ KPI to report a valid key figure in the company.

Further research could examine whether the UEQ KPI can be implemented in an organization as a KPI. Emphasis should be placed on interpretability and acceptance. In addition, it could determine whether the UEQ KPI meets all the requirements for a KPI according to the SMART model [44].

We used Amazon and Skype in our study. However, these products tend to have a high proportion of pragmatic quality (Fig. 1 and 2). Besides, in further research, we have to find out what influence products with rather high hedonic quality have on the UEQ KPI.

In addition, it should be investigated to what extent cultural differences are reflected in the KPI. The results show that KPIs reflect cultural differences, but not how strongly and for what reason.

7. References

- [1] M. Hodges-Schell, J. O'Brien, *Communicating the UX vision: 13 anti-patterns that block good ideas*, Morgan Kaufmann, Waltham, MA, 2015.
- [2] G. A. Boy, *The Handbook of Human-Machine Interaction: A Human-Centered Design Approach*, 1st Edition, CRC Press, Milton, 2017.
- [3] M. Hassenzahl, The effect of perceived hedonic quality on product appealingness, *INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION* 13(4) (2001) 481–499.
- [4] J. Preece, Y. Rogers, H. Sharp, *Interaction design: Beyond human-computer interaction*, 4th Edition, Wiley, Chichester, 2015.
- [5] DIN, *Ergonomics of human-system interaction - part 210: Human-centred design for interactive systems* (2010).
- [6] D. A. Norman, *Emotional Design: Why We Love (or Hate) Everyday Things*, Basic Books, New York, 2007.
- [7] N. Tractinsky, Aesthetics and apparent usability, in: S. Pemberton (Ed.), the SIGCHI conference, 1997, pp. 115–122. doi:10.1145/258549.258626.
- [8] C. E. Osgood, G. J. Suci, P. H. Tannenbaum, *The measurement of meaning*, University of Illinois Press, Urbana-Champaign, 1978.
- [9] B. Laugwitz, T. Held, M. Schrepp, Construction and evaluation of a user experience questionnaire, in: A. Holzinger (Ed.), *HCI and Usability for Education and Work*, Vol. 5298 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 63–76. doi:10.1007/978-3-540-89350-9_6.
- [10] M. Rauschenberger, M. Schrepp, M. Perez-Cota, S. Olschner, J. Thomaschewski, Efficient measurement of the user experience of interactive products. how to use the user experience questionnaire (ueq).example: Spanish language version, *International Journal of Interactive Multimedia and Artificial Intelligence* 2 (1) (2013) 39. doi:10.9781/ijimai.2013.215.
- [11] M. P. Cota, J. Thomaschewski, M. Schrepp, R. Gonçalves, Efficient measurement of the user experience. a portuguese version, *Procedia Computer Science* 27 (2014) 491–498. doi:10.1016/j.procs.2014.02.053.
- [12] R. Alves, P. Valente, N. J. Nunes, The state of user experience evaluation practice, in: V. Roto, T. Olsson, K. Väänänen-Vainio-Mattila, J. Häkkinen, E. Hvannberg, O. Juhlin (Eds.), *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, ACM Press, New York, New York, USA, 2014, pp. 93–102. doi:10.1145/2639189.2641208.
- [13] N. Rasmussen, C. Y. Chen, M. Bansal, *Business dashboards: A visual catalog for design and deployment*, John Wiley & Sons, Hoboken, N.J, 2009.
- [14] National Institute of Standards and Technology, *Performance measurement guide for information security* (July 2008).
- [15] J. Lazar, J. H. Feng, H. Hochheiser, *Research methods in human-computer interaction*, Wiley, Chichester and West Sussex and U.K, 2010.
- [16] J. Brooke, Sus: A quick and dirty usability scale, in: P. W. Jordan, B. Weerdmeester, A. Thomas, I. L. Mclelland (Eds.), *Usability evaluation in industry*, Taylor and Francis, London, 1986.
- [17] J. Kirakowski, M. Corbett, Sumi: The software usability measurement inventory, *British Journal of Educational Technology* 24 (3) (1993) 210–212. doi:10.1111/j.1467-8535.1993.tb00076.x.
- [18] M. Moshagen, M. T. Thielsch, Facets of visual aesthetics, *International journal of human-computer studies* 68 (10) (2010) 689–709. doi:10.1016/j.ijhcs.2010.05.006.
- [19] J. Sauro, Supr-q: A comprehensive measure of the quality of the website user experience, *Journal of Usability Studies* 2015 (10) (2015) 68–86.
- [20] J. R. Lewis, Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, *International Journal of Human-Computer Interaction* 7 (1) (1995) 57–78. doi:10.1080/10447319509526110.
- [21] F. F. Reichheld, The one number you need to grow, in: *Harvard Business Review*, 2003, pp. 47–54.
- [22] J. Sauro, J. S. Dumas, Comparison of three one-question, post-task usability questionnaires, in: D. R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, S. Greenberg (Eds.), *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press, New York, New York, USA, 2009, p. 1599. doi:10.1145/1518701.1518946.
- [23] J. R. Lewis, Psychometric evaluation of the psuq using data from five years of usability studies, *International Journal of Human-Computer Interaction* 14 (3-4) (2002) 463–488. doi:10.1080/10447318.2002.9669130.
- [24] J. Sauro, E. Kindlund, A method to standardize usability metrics into a single score, in: G. van der Veer, C. Gale (Eds.), *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, ACM Press, New York, New York,

- USA, 2005, p. 401. doi:10.1145/1054972.1055028.
- [25] M. Schrepp, M. Pérez Cota, R. Gonçalves, A. Hinderks, J. Thomaschewski, Adaption of user experience questionnaires for different user groups, *Universal Access in the Information Society* doi:10.1007/s10209-016-0485-9.
- [26] H. Santoso, M. Schrepp, A. Hinderks, J. Thomaschewski, Cultural differences in the perception of user experience, in: *Gesellschaft für Informatik (Ed.), Mensch und Computer 2017*, 2017. doi:10.18420/muc2017-mci-0272.
- [27] M. Schrepp, *User experience questionnaire handbook: All you need to know to apply the ueq successfully in your projects (2015)*.
URL <http://www.ueq.online.org>
- [28] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, J. Thomaschewski, Ueq kpi result report of the evaluation of amazon and skype. doi:10.13140/RG.2.2.34239.76967.
- [29] J. C. Nunnally, I. H. Bernstein, *Psychometric theory*, 3rd Edition, McGraw-Hill higher education, Tata McGraw Hill Education Private Ltd, New Delhi, 2010.
- [30] L. J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* 16 (3) (1951) 297–334. doi:10.1007/BF02310555.
- [31] T. K. Landauer, K. M. Galotti, S. Hartwell, Natural command names and initial learning: A study of text-editing terms, *Communications of the ACM* 26 (7) (1983) 495–503. doi:10.1145/358150.358157.
- [32] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Elsevier Science, Burlington, 2013.
- [33] M. Schrepp, A. Hinderks, J. Thomaschewski, Construction of a benchmark for the user experience questionnaire (ueq), *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (4) (2017) 40–44. doi:10.9781/ijimai.2017.445.
- [34] M. Hassenzahl, N. Tractinsky, User experience – a research agenda, *Behaviour & Information Technology* (25) (2006) 91–97.
- [35] H. C. Triandis, Individualism-collectivism and personality, *Journal of Personality* 69 (6) (2001) 907–924. doi:10.1111/1467-6494.696169.
- [36] T. Parsons, B. S. Turner, *The social system*, 2nd Edition, Routledge sociology classics, Routledge, Abingdon, Oxon, 2005.
- [37] S. H. Schwartz, Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries, in: *Advances in Experimental Social Psychology Volume 25, Vol. 25 of Advances in Experimental Social Psychology*, Elsevier, 1992, pp. 1–65. doi:10.1016/S0065-2601(08)60281-6.
- [38] S. H. Schwartz, A. Bardi, Value hierarchies across cultures, *Journal of Cross-Cultural Psychology* 32 (3) (2016) 268–290. doi:10.1177/0022022101032003002.
- [39] H.-E. Hoffmann (Ed.), *Internationales Projektmanagement: Interkulturelle Zusammenarbeit in der Praxis, orig.-ausg Edition*, Vol. 50883 of Dtv, Dt. Taschenbuch-Verl., München, 2004.
- [40] A. Hinderks, M. Schrepp, J. Thomaschewski, A benchmark for the short version of the user experience questionnaire, in: M. J. Escalona, F. J. Domínguez Mayo, T. A. Majchrzak, V. Monfort (Eds.), *WEBIST 2018, SCITEPRESS – Science and Technology Publications, Lda.*, 2018, pp. 373–377.
- [41] J. Winter, K. Rönkkö, M. Rissanen, Identifying organizational barriers—a case study of usability work when developing software in the automation industry, *Journal of Systems and Software* 88 (2014) 54–73. doi:10.1016/j.jss.2013.09.019.
- [42] D. Mian, M. Humphreys, A. Sidwell (Eds.), *Construction projects immediate health check: a CSF & KPI approach*, 2004.
- [43] E. Roubtsova, V. Michell, Modelling and validation of kpis, in: *Proceedings of the Third International Symposium on Proceedings of the Third International Symposium on Business Modeling and Software Design, SCITEPRESS - Science and Technology Publications*, 2013, pp. 96–105. doi:10.5220/0004774400960105.
- [44] G. T. Doran, There's a s.m.a.r.t. way to write management's goals and objectives, in: *Management Review*, Vol. 11, 1981, pp. 35–36.