Selecting the Appropriate User Experience Questionnaire and Guidance for Interpretation: the UEQ Family

Jessica Kollmorgen[1], Andreas Hinderks[2], Jörg Thomaschewski[1]*

[1] University of Applied Sciences Emden/Leer (Germany)
[2] University of Applied Sciences and Arts Hannover (Germany)

* Corresponding author: joerg.thomaschewski@hs-emden-leer.de

Received 6 June 2024 | Accepted 15 July 2024 | Early Access 22 August 2024

Abstract

Measuring the user experience (UX) of products, systems and services is individual depending on the research question. On the one hand, the user's goals and environment play a role in the subjective evaluation. On the other hand, different UX factors are relevant depending on the product. In this case, it is practical to have a questionnaire family as an aid, whose questionnaires are geared towards these different use cases. The User Experience Questionnaire (UEQ) family allows researchers and practitioners to choose the right tool for efficient UX measurement from three questionnaire versions. This article summarizes the UEQ, its short version (UEQ-S) and a modular framework (UEQ+) with overall 27 UX factors and purposes in over 30 different languages. In addition, specific instructions and assistance are provided for the statistical evaluation and interpretation of the questionnaire results. With the help of a key performance indicator (KPI), benchmarks and an importance-performance analysis (IPA), the realization of UX measurements is made easier for researchers and practitioners. To make it even more convenient to choose the right questionnaire from the UEQ family, influencing factors on the UX measurement and recommendations for action are given.

Keywords

Agile Methods, UEQ, UEQ+, UEQ-S, User Experience, UX, UX Measurement, User Experience Questionnaire, UX Management.

DOI: 10.9781/ijimai.2024.08.005

I. Introduction

Researchers and practitioners have been dealing with the integration of actual user needs into the development of products, systems, and services for years to avoid late and often costly changes. Similar to agile development, it is necessary to continuously and frequently gather feedback from users. This forms the basis of the user experience research field, which focuses on planning, measuring, and evaluating user feedback. That is done not only to ensure efficiency and effectiveness but also to address specific emotions and subjective beliefs such as fun of use or aesthetics before, during and after the use [1].

In practice, the question arises of how to make these subjective impressions tangible in order to set concrete goals for improvements, or even the validation of achieved goals. Qualitative measurement methods offer one approach to do this, which involves among other things evaluating specific designs or individual cases for interpretive analysis. User interviews, for example, are well-suited for this qualitative analysis, generating some relevant insights with just a few participants. These measurement methods answer the "why" question, but can be time-consuming and not always generalizable to specific product or company goals [2].

In contrast, quantitative measurement methods provide another approach, using statistical data to analyze and identify concrete trends [3]. These methods often serve as the basis for defining milestones in practice. However, acquiring large amounts of these quantitative data can be a lenghty process. Related work has shown that interpretations should not be made about populations based on small samples [4], [5], as they are more susceptible to disturbances and external influences. At the same time, the measurement method must not be too burdensome to allow for participation within a realistic timeframe, such as within the scope of daily life. A fundamental goal of user experience research is therefore to structure and simplify the capture of subjective perceptions.

An established measurement tool for quantitative research, minimizing the effort of recording and participating, is user experience questionnaires [6]. These enable the capture of subjective impressions, opinions, and evaluations from users regarding various aspects of the user experience. This makes it possible to generate large amounts of statistical data that can be further used for the analysis of products, systems, and services. User experience factors such as functionalities (e.g., easy to use, understandable) or sensations (e.g., entertaining) can be captured with an appropriate time investment from participants [7].

An established questionnaire family in this regard is the long-established User Experience Questionnaire (UEQ) [8], which has been continuously analyzed and expanded since 2008. This questionnaire defines various validated factors that measure specific aspects of

Please cite this article as:
J. Kollmorgen, A. Hinderks, J. Thomaschewski. Selecting the Appropriate User Experience Questionnaire and Guidance for Interpretation: the UEQ Family, International Journal of Interactive Multimedia and Artificial Intelligence, (2024), http://dx.doi.org/10.9781/ijimai.2024.08.005

user experience. However, as explained, the goal of user experience research is to simplify the answering of these questionnaires as much as possible, which is why a short version of the questionnaire (UEQ-S) [7] exists since 2017. It should be noted that not all identified UX aspects are suitable or relevant for every industry or product, which is why a modular version of the questionnaire (UEQ+) [9] was published in 2019. This modular version can be assembled for individual use cases or research questions. This questionnaire family thus allows for a comprehensive quantification of subjective data.

In this article, we assist researchers and practitioners in selecting the appropriate questionnaires and user experience factors for the individual use cases and research questions. In addition, we provide specific assistance for the statistical analysis and interpretation of the questionnaire results. An overview of the UEQ Family including its components and their relations is shown in Fig. 1.

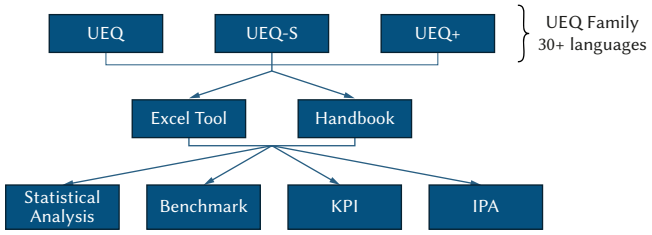

Fig. 1. Overview of the UEQ Family and its components, including the statistical analysis, benchmark, key performance indicator (KPI) and importance-performance analysis (IPA).

The structure of the article is therefore as follows: after an overview in Section II of basic concepts and related work on measuring the expected and perceived user experience with the help of questionnaires, Section III will first present the specific foundations and differences of the individual questionnaires of the UEQ family, including their measured factors and applicability, also in the context of other questionnaires. Subsequently, in Section IV, an insight into the interpretation quantitative UX measurement will be provided to researchers and practitioners on the example of the UEQ family, including relevant statistical analyses using methods such as the UEQ Tool, benchmarks, relevant key performance indicators, and the importance-performance analysis. These reference points and comparison values help to interpret and practically classify the measured results. Section V will focus on relevant user- and product-related influencing factors that can affect both the subjective perceived user experience and the interpretations of the results, which should be considered. Finally, this forms the basis for concrete recommendations for action and an outlook in Section VI.

II. Basic Concepts

This section introduces and explains some basic concepts that are necessary for the general argumentation and interpretation of this article, including the differentiation between usability and user experience and the measurement of user experience using questionnaires.

A. Usability and User Experience

If we imagine products, systems or services, such as online banking, then the users' expected requirements focus on tasks that they want to achieve with them. Factors such as efficiency and effectiveness are relevant for operation. This is associated with the concept of usability, which according to the established ISO 9241-11 is described as "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments." [10]

However, if we now take a look at modern social networks, factors have less to do with the fulfillment of tasks and more to do with individual perceptions, such as aesthetics or novelty. The concept of user experience (UX) thus expands the original usability concept with additional factors in order to provide a more holistic overview of the interaction between user and product[1]. User experience can therefore be described as the "user's perceptions and responses that result from the use and/or anticipated use of a system, product or service" [11], including for example also emotions or beliefs.

Since the 1990s, there have been various models that attempt to describe usability and user experience as a whole. Each model has a different focus for mapping UX with the help of factors [1]. For example, Hassenzahl's [12] model takes a holistic approach to UX, which describes that UX consists of different dimensions: *pragmatic* and *hedonic* quality. While the *pragmatic* UX aspects relate to the usability and functionality of a product in a psychological context, the *hedonic* UX aspects relate more to emotions, such as the experience of pleasure or aesthetics. This theoretical construct also has a practical background, as could be also shown in a study from 2023 [13]. Here, it was shown that products with a differently weighted pragmatic or hedonic focus were also rated differently.

However, as the subjectivity of the aspects already makes clear, different users or target groups may judge the same product differently in terms of its UX. This may be due to the fact that they have different needs or abilities when using the product. For example, older users might experience social networks quite differently to younger, more tech-savvy users [14]. According to Hassenzahl (2003) [12], the different usage modes also play a role in the subjectivity of the evaluation. While the *goal mode* is aimed at the pure fulfillment of goals, in *action mode* the goal is not fixed in advance, but arises during use. In the latter case, the use of the product itself can also be the goal, e.g. arising from boredom.

It is therefore necessary to check the expected and perceived UX of products not just at one point in time, but continuously. This can refer to the time before (anticipated UX), during (momentary UX), after the use (episodic UX) and to the course over time (cumulative UX) [15]. To be able to examine this, the UX must first be made measurable [1].

B. UX Measurement

Users expect a high level of satisfaction when interacting with the product, both for simple and complex products that specialize in functionality rather than user satisfaction. They are expected to be able to use the product efficiently without much effort [1]. But how exactly can individual and subjective UX be continuously measured and how can generalized conclusions be drawn?

There are various evaluation approaches to prevent, for example, the design from mismatching with the user's mental model [16]. Well-known methods for this are, for example, user interviews or heuristic evaluation [17]. Relevant findings of these methods can ultimately be drawn to improve the UX of products. In order to implement changes in a standardized way without losing sight of other aspects that are already good, researchers and experts developed so-called usability and UX heuristics. These compromise guidelines for identifying and eliminating potential UX weaknesses at an early stage [18].

While these qualitative methods can also capture complex views, they are time-consuming and resource-intensive. They reach their limits when it comes to the scalability of UX measurement for many users and different target groups [14]. For example, it should be avoided that certain user groups, such as older or less educated people or people with disabilities, are excluded from use because they are not considered and thus experience a poor user experience [19].

[1] In the rest of the article, "product" is used as a substitute for "product, system or service".

In this case, quantitative user experience questionnaires are a suitable tool. Thanks to their simple scalability and comparability as well as efficiency and standardization in data collection, they allow a wide range of aspects to be surveyed and flexibly adapted to the target group without generating high costs. They can measure the expected and perceived UX of products and form a basis for improvements to the product under investigation [7].

C. UX Questionnaires

In recent decades, various standardized questionnaires have been established that measure both usability and user experience factors. A factor is equated with an actual and real UX aspect, and consists of various individual items that are assigned to this factor in advance using suitable methods such as statistical factor analysis. A standardized questionnaire therefore contains multiple UX factors, which in turn contain different items [1].

Various forms of items can be used when dealing with questionnaires. Commonly used are 5- or 7-point Likert scales, which users can use to express their level of (dis)agreement in relation to a short statement [20]. One example of this is the SUS (System Usability Scale) questionnaire, which is shown in Fig. 2.

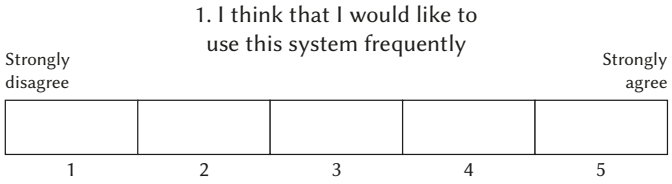

Fig. 2. Example of an item from the System Usability Scale (SUS) [21].

However, the disadvantage of this item format is the scope for interpretation. Therefore, there are also semantic differentials that use certain clear, contradictory terms in relation to a UX factor. This is the case with the UEQ (User Experience Questionnaire), an example of which is shown in Fig. 3.

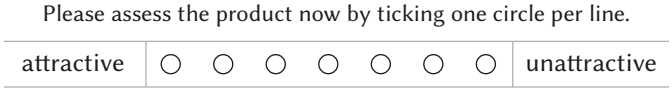

Fig. 3. Example of an item from the User Experience Questionnaire (UEQ) [22].

These contrasting attributes can reduce misunderstandings and inconsistencies in interpretation [20].

With the help of holistic questionnaires, it is therefore possible to determine the satisfaction of various relevant user groups with individual products and use cases. However, UX questionnaires differ in their area of application and focus. While the AttrakDiff [23], for example, lays a greater emphasis on hedonic than on pragmatic quality aspects, which can be not entirely suitable for the assessment of professional software [8], the UEQ [8] and the meCUE [24] are broader in scope and measure both the usability and the user experience for products in a wide range of application areas. Based on their regular use, it is possible to continuously record the expectations and perceptions of users and improve the quality of the UX on this basis [1].

III. Fundamentals

As explained in Section II, UX questionnaires enable a quick but also comprehensive measurement and evaluation of the expected and perceived UX of products for larger user target groups. For example, users can be presented with a short questionnaire when they leave a web service to enable data collection from large samples even with little effort [14]. These questionnaires can be provided both online and in printed form.

Different questionnaires take various UX factors into account and are effective for their specific contexts. The UEQ Family is also designed for different contexts and offers a flexible approach. It is designed to be highly adaptable, allowing for customization to suit a wide range of use cases in UX measurement. This adaptability ensures the provision of relevant and accurate insights across different contexts and user groups.

The individual products of the UEQ family, the UEQ, UEQ-S (short version) and UEQ+ (modular version), are presented below with the respective descriptions of UX factors, items and intended use. The UEQ family is currently available in over 30 languages.

A. UEQ[2]

The User Experience Questionnaire was published in 2008 by Laugwitz et al. [8] to quantitatively measure the perceived UX of products with their respective functions. The UEQ follows three main objectives. Firstly, the application and analysis of the questionnaire should enable a **quick assessment** as far as possible. Secondly, with the help of end users, a **comprehensive impression of user experience** of the product under investigation should be gained. Thirdly, users should be given the opportunity to express feelings/impressions that arise when experiencing the product in a **simple and immediate** way as possible. In 2008, other available questionnaires also focussed on one or two main objectives, but in no case all three [8].

Fig. 4. Overview of the UEQ and the included UX factors with their meaning and corresponding items.

[2] available online free of charge under https://www.ueq-online.org/

The UEQ measures a total of six UX factors (see Fig. 4), which can be assigned to *pragmatic* and *hedonic* quality aspects as well as the attractiveness of the product. Each of the factors is measured using specific items. While the factor *attractiveness* is measured with six items, the other five factors are determined using four items each. As a result, the UEQ with the six UX factors contains a total of 26 items, which take a total of 3-5 minutes to answer [7].

The process of selecting the items was extensive [8]. In total, a set of originally more than 200 potential UX items was created in a study with 153 participants. In the next step, this selection was shortened with a number of experts to a rough version of 80 items. A statistical factorial analysis was used to check which of these items could be assigned to which of the six factors as clearly as possible and free from misinterpretation. After completion of the UEQ, the questionnaire was additionally validated in two studies [8].

Semantic differentials were chosen as the item format, which are recorded on a 7-point Likert scale with a range from -3 (worst) to +3 (best). Half of the positive attributes are on the right-hand side, the other half on the left-hand side of the questionnaire. This is to avoid the introduction of a response tendency with a one-sided item polarity [8], [25]. The distribution of the item order is randomized [7], [8]. An excerpt of the different items can be seen using the example of the factor *Efficiency* in Fig. 5.

To achieve my goals, I consider the product as

slow	○	○	○	○	○	○	○	fast
inefficient	○	○	○	○	○	○	○	efficient
impractical	○	○	○	○	○	○	○	practical
cluttered	○	○	○	○	○	○	○	organized

Fig. 5. Items of the factor *Efficiency* measured with the UEQ [22].

In 2019, the UEQ was expanded to include a key performance indicator (KPI) [26]. This was due to the desire of decision-makers to interpret single numerical values. To obtain an UX KPI, it was first necessary to measure the relative importance of the individual six UEQ factors in relation to the product under investigation. This means that an additional question on the importance of the specific factor was added to each of the six factors, which can be rated on a 7-point scale. An example of the factor Efficiency is shown in Fig. 6.

I can perform my tasks with the product fast, efficient and in a pragmatic way. The user interface looks organized.

Not important at all	○	○	○	○	○	○	○	Very important

Fig. 6. Item related to the key performance indicator (KPI) of the factor *Efficiency* measured with the UEQ [22].

These importance ratings are used to determine additional UX information (see Section IV.C).

At the beginning of further development (2008), the UEQ was originally available in German and English. However, in order to enable the most international use possible in a wide range of industries and use cases, the UEQ is now available for download in 36 languages, including Spanish, Finnish, Russian, Chinese or Turkish [27]. It is also available in a simpler version with simplified language, e.g. for children or people with disabilities [28]. Over time, the versions created have been repeatedly checked and improved, further reducing the risk of misinterpretation [29].

In order to avoid misunderstandings when implementing and evaluating the UEQ, a website [27], a handbook [22] and a supporting Excel tool [30] are available to facilitate statistical analysis, so only the individual data need to be inserted.

This makes the UEQ an easy-to-use, reliable and valid method for quickly capturing the user experience of products within a few minutes, and is short enough to be completed both in printed and online form, even if demographic questions are added at the beginning or end of the questionnaire to gain an additional impression of the target group [14]. It can also supplement or support data from other evaluation methods with its subjective quality ratings [7].

B. UEQ-S[3]

Even if it only takes 3-5 minutes to complete the UEQ, there are situations in which this time is not available, for example if the user to be surveyed leaves a web service. Other possible situations would be that the respondents are to evaluate several product (variants) in one go, or the questionnaire is to be integrated into existing product experience questionnaires, which would exceed the reasonable total length. In these and other cases, it is helpful to be able to fall back on a short version.

For this reason, a short version of the UEQ, the UEQ-Short (UEQ-S), was published in 2017 by Schrepp et al. [7]. In the UEQ-S, the measurement of the six single dimensions like *Efficiency* were skipped, and only the meta-dimensions *Pragmatic Quality* and *Hedonic Quality* are measured. Instead of 26 items of the UEQ, these two dimensions are measured with a total of only 8 items, which are shown in Fig. 7 (4 items of the *pragmatic* factors as well as 4 items of the *hedonic* factors which fit best from the UEQ). The mean value of the 8 items is evaluated as the overall UX value [7].

The positive attributes of the semantic differentials are on the same side, as the focus of the short version is on the required completion time and it was assumed that a one-sided item polarity reduces the completion time of the questionnaire. Furthermore, the order of the items is not randomized: the first 4 items reflect the *pragmatic* quality, the other 4 items, the *hedonic* quality.

obstructive	○	○	○	○	○	○	○	supportive
complicated	○	○	○	○	○	○	○	easy
inefficient	○	○	○	○	○	○	○	efficient
confusing	○	○	○	○	○	○	○	clear
boring	○	○	○	○	○	○	○	exciting
not interesting	○	○	○	○	○	○	○	interesting
conventional	○	○	○	○	○	○	○	inventive
usual	○	○	○	○	○	○	○	leading edge

Fig. 7. Items of the short version of the UEQ (UEQ-S) [27].

A data set with 1867 data records was used as the basis for creating the UEQ-S, in which a total of 21 different products from different areas of application were assessed (e.g. webshops or business software) [7]. Based on this data, a main component analysis was carried out on all items of the *pragmatic* factors and then on the items of the *hedonic* factors. In order to prove that the scale is precise and that items do not overlap with other items, the 8 items with the lowest factor loadings according to the analysis were selected for the short version. The constructed UEQ-S was also evaluated in a further study with 47 people and 3 products. A high level of consistency was shown and the previously measured results were confirmed, meaning that the questionnaire is consistent and stable in practice [7].

[3] available online free of charge under https://www.ueq-online.org/

The mean values of the four items were then compared with the mean values of the respective 16 and 12 items of the *pragmatic* and *hedonic* factors. It was found that the differences were close to zero, meaning that the short version is able to accurately predict the values of the full UEQ. As the UEQ-S is therefore a subset of the UEQ, this also means that the UEQ-S can be used in all 30+ languages in which the UEQ is available [7]. The short version provides an approximate assessment of the UX quality of a product based on higher level meta-dimensions and is not intended to replace the UEQ. It is therefore recommended to measure the six UX factors in detail if the results are to be interpreted for improvement potential. In such cases, e.g., when used together with usability tests, the small increase in efficiency in the completion time does not compensate for the loss of detailed impressions of the quality aspects.

C. UEQ+[4]

As the user experience is a subjective construct of various factors, it follows that not all of these UX quality aspects are actually relevant depending on the product and use case. Experience with the UEQ has shown that UX factors are missing (e.g. Trust [31]) and factors available in the UEQ are sometimes not needed (e.g. Stimulation [32]).

These considerations gave rise to the UEQ+ (UEQplus) questionnaire, which was published in 2019 by Schrepp and Thomaschewski [9]. This is a modular framework that allows researchers and practitioners to select the relevant UX factors for the desired use case from a catalog of UX factors.

Since the factors can be combined individually depending on the use of the questionnaire, it was necessary to change the randomized format of the semantic differentials to a one- sided item polarity in which all positive terms are positioned on the right-hand side. Since the order of the selected factors should also be able to be determined individually, all items of a factor were grouped and placed in context to prevent misinterpretation. For this purpose, a sentence was added above the items of a factor (see Fig. 8). However, the 7-point Likert scales were retained as their use in the UEQ has been empirically validated. All factors therefore have the same response format so that they can be easily combined [9].

In order to be able to make a judgment as to whether the selected factors are actually relevant for the use case or the product, an importance rating was also added to each UX factor using a 7-point Likert scale (see Fig. 8), similar to the KPI extension of the UEQ.

To achieve my goals, I consider the product as								
slow	○	○	○	○	○	○	○	fast
inefficient	○	○	○	○	○	○	○	efficient
impractical	○	○	○	○	○	○	○	practical
cluttered	○	○	○	○	○	○	○	organized
I consider the product property described by these terms as								
Completely irrelevant	○	○	○	○	○	○	○	Very important

Fig. 8. Example of the UX factor *Efficiency* and the corresponding items [5].

To add further factors to the six UX factors of the UEQ, an empirical study was conducted in 2018 with 192 participants, the results of which were examined in a principal component analysis. Based on this, four items with the highest factor loadings per factor were used for newly created factors. The study resulted in the following UX factors, which can be measured using the UEQ+ [9], consisting of the factors of the UEQ as well as new factors:

- **Attractiveness:** Overall impression concerning the product. Do users like or dislike it?[5]
- **Perspicuity:** Impression that it is easy to learn how to use the product.
- **Efficiency:** Impression that tasks can be finished without unnecessary effort.
- **Dependability:** Impression to be in control of the interaction with the product.
- **Stimulation:** Impression that it is interesting and fun to use the product.
- **Novelty:** Impression that the product design or product idea is creative and original. [33]
- **Trust:** Subjective impression that the data entered into the product are in safe hands and are not used to the detriment of the user.
- **Aesthetics:** Impression that the product looks nice and appealing.
- **Adaptability:** Subjective impression that the product can be easily adapted to personal preferences or personal working styles.
- **Usefulness:** Subjective impression that using the product brings advantages, saves time or improves personal productivity.
- **Intuitive Use:** Subjective impression that the product can be used immediately without any training, instructions or help from other persons.
- **Value:** Subjective impression that the product is of high quality and professionally designed.
- **Trustworthiness of Content:** Subjective impression that the information provided by the product is reliable and accurate.
- **Quality of Content:** Subjective impression that the information provided by the product is up to date, well-prepared and interesting. [9]

Factors that are specifically geared towards household appliances (e.g. washing machines) were also included:

- **Haptics:** Subjective feelings resulting from touching the product.
- **Acoustics:** Subjective impression concerning the sound or operating noise of the product. [9]

These factors with the associated items were then evaluated in a further study using three product categories (webshops, video platforms, programming environments) and two products each. A high factor quality was proven in this study. In addition, the selected factors for each product were considered important for the participants [9].

In further studies after 2019, additional UX factors were also included in the factor catalog. For example, the UEQ+ can now measure UX factors of voice assistants (e.g. Siri or Alexa):

- **Response behavior:** Impression that the voice assistant behaves respectful and trustworthy.
- **Response quality:** Impression that the responses of a voice assistant cover the user's information needs.
- **Comprehensibility:** Impression that the voice assistant correctly understands the users instructions and questions using natural language. [34]

The UX of complex medical devices (e.g. MRI and CT scanners) can also be measured:

- **Result quality:** Can goals and results be fully and accurately achieved by using the product?

[4] available online free of charge under https://ueqplus.ueq-research.org/

[5] In contrast to the UEQ, the factor *Attractiveness* is measured with 4 items in the UEQ+.

- **Hardware security:** Does the hardware bear risks, which might be hazardous to health?
- **Risk handling:** Can users identify and handle risks and errors? [35]

And factors for the influence of using or owning a product on social connections/status can also be measured:

- **Identification:** Impression that using or owning a product influences the social status.
- **Social interaction:**Impression of the user that the product supports social activities or helps to build social contacts.
- **Social stimulation:** Impression of the user regarding the anticipated social gains resulting from their interaction with a product.
- **Social acceptance:** Impression of the user regarding how they are accepted and approved by others and themselves when using a product. [36]

Over the years, the UEQ+ has been developed as a modular questionnaire that allows researchers and practitioners to measure a total of 27 UX factors depending on the research question. The factors, which all have the same item and response format, can be combined individually and are easy to answer.

However, there are also some additional expenses compared to the original UEQ. The effort required to compile and evaluate the modular questionnaire is higher, as it is not always known which factors need to be selected for each product. However, we will present impact factors and recommendations for action below to make the choice easier. In addition, a handbook [22] and a supporting Excel tool [37] are also offered for the UEQ+, to facilitate statistical analysis, so only the individual data need to be inserted. Benchmarks are also provided, for example, to make it easier to interpret the results. The UEQ+ is also currently available in 25 languages [33].

D. Differentiation From Other Established Questionnaires

As already explained in Section II, there are a large number of UX factors and corresponding questionnaires that measure these factors. This section therefore aims to show the relationship between the UEQ family and established related questionnaires. For more in-depth analyses, corresponding studies are recommended [1], [20].

In a comparison of the German-language version of the **UEQ** from 2018 [38] with the German-language questionnaires VISAWI (Visual Aesthetics of Website Inventory) and meCUE, three products with different usage contexts were selected. While the VISAWI measures factors such as *Variety* or *Colorfulness*, the meCUE measures three modules based on the "Components of User Experience" model: Product perception (factors e.g. *Usefulness*, *Usability* or *Visual Aesthetics*), user emotions (positive and negative) and consequences (factors *Product loyalty* or *Intention to use*). On the one hand, there were high correlations, e.g. between the *hedonic* UEQ factors and the VISAWI factors, which was to be expected according to the underlying psychological model of the questionnaires. On the other hand, however, the correlations of the questionnaire factors varied depending on the product, which means that other influences such as the context of use may be present here.

In another comparison of the questionnaire results of the **UEQ-S** with SUS and UMUX-LITE, a study [13] with four products showed that all three questionnaires delivered almost identical results in the evaluation of the UX quality of the products. This was also evident in the scale correlations, as UMUX-LITE and SUS showed high correlations with each other. This confirmed the suggestion [39], [40] that the results of the UMUX-LITE can predict those of the SUS in scenarios with a low number of questions. On the other hand, the dimension of *pragmatic* quality of the UEQ-S also showed high to very high correlations with the SUS and UMUX-LITE. However, there were differences in the correlations with the *hedonic* dimension, which was to be expected as usability questionnaires are aimed at *pragmatic* quality.

In 2020 [41], the correlations between five UX factors of the **UEQ+**, the complete SUS and the Net Promoter Score (NPS) were also examined. The NPS is used to measure customer loyalty and is not a classic usability questionnaire, but suggests that poorly usable products would not maintain a high NPS. The factors *Intuitive Use*, *Quality of content*, *Trustworthiness of content*, *Trust* and *Stimulation* were selected as the equivalent of the UEQ+ for the comparison of the evaluation of a product. The key performance indicators of the questionnaires were calculated on the basis of the results. It was found that the NPS correlates positively and linearly with the SUS and UEQ+ KPI moderator. The SUS and UEQ+ KPI also correlate strongly with each other in a positive linear fashion, especially the UX factor *Intuitive use*.

In summary, it can therefore be said that similarities between the *pragmatic* factors of the UEQ, UEQ-S and UEQ+ and classic usability questionnaires have already been identified in the past. The UEQ family can therefore be used as a replacement or supplement to comparable established usability questionnaires (e.g., SUS, UMUX, ISONORM). In addition, based on a more holistic model of user experience, it measures UX factors that cannot be measured completely and/or in several languages by other UX questionnaires (e.g., AttrakDiff2, VISAWI, meCUE).

IV. Interpreting Results From UX Questionnaires

How is it that the same product is sometimes perceived very differently by separate individuals? On the one hand, general influences such as demographic factors (e.g., age or gender) or cultural background matter. The different levels of experience already gained with the product can also have an influence.

On the other hand, the significance of the measured values is tangible if physical properties of objects are to be measured, such as the weight of an object or the reaction time of a system to an input. However, the situation is different with psychological properties, such as aesthetics or stimulation. These concepts need to be measured just as carefully in order to make different individual opinions tangible.

For this reason, the recommended statistical instruments of the UEQ Family are presented below, which are necessary to make the different results both comparable and, above all, interpretable.

A. Statistical Analysis

Once the subjective UX has been recorded using the questionnaires, it is necessary to evaluate and interpret the results. To make it easier to get started, Excel sheets are provided for the UEQ family, depending on the questionnaire version used (UEQ [30], UEQ-S [42], UEQ+ [37]). The raw data, i.e. the cell-by-cell scores between -3 and +3, are inserted into this sheet. The tool then first returns the mean values per participant, item and factor, with the corresponding position (standard deviation) and spread (variance). Based on this, also the confidence interval is calculated in the tool. The factor means and confidence intervals are shown as an example in Fig. 9.

It is clear that every factor within the range from -3 to 3 was rated higher than 0.7. This means that the overall UX of the product was perceived as rather positive (green area). Nevertheless, the *pragmatic* factors *Perspicuity* and *Efficiency* were rated best for this fictitious product, while the *hedonic* factors *Stimulation* and *Novelty* were rated worst and are still partially in the neutral range (yellow). On the one hand, these are indications on which product features require improvement. On the other hand, the use case of the product should be considered at this point. Products with a focus on achieving *pragmatic*

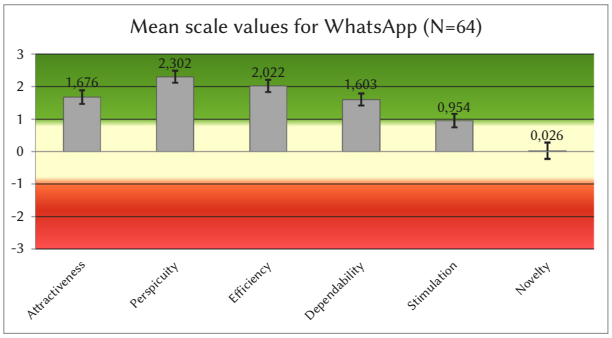

Fig. 9. Product evaluation for WhatsApp (N=64) [25], measured with the UEQ and evaluated with the UEQ Excel tool [30].

goals (e.g., word processing with Microsoft Word) should, as expected, score better on the *pragmatic* factors, while leisure products with a focus on entertainment (e.g., social networks such as TikTok) should expectedly score better on *hedonic* factors. The confidence intervals (lines at the edge of the bars) in Fig. 9 are small (< 0.25). Since not all actual users can be reached in the UX measurement, but the surveys are carried out with the help of random samples, this assessment is necessary in order to draw conclusions about the range in which the actual factor mean lies. The smaller the line, the more likely a similar result is for other samples. If, on the other hand, the lines of the confidence interval are wide, the results should not be overinterpreted. This can occur in the case of small sample sizes or several outliers (e.g. a few very good or very poor ratings by individual participants), making further investigations necessary. In this case, for example, the examination of inconsistencies in the Excel tools can be used, which provides information on whether some answers may not have been answered seriously by participants. This is the case if several items on the same factor were answered very differently (more than 3 points different) (see Fig. 10).

Scales with inconsistent answers		
Pragmatic Quality	Hedonic Quality	Critical?
1		1
		0
1		1
		0
		0
		0
		0
1	1	2

Critical length	
Same answer for	Middle Category
2	
4	
3	
3	
5	
8	Remove
8	
3	

Fig. 10. Example of inconsistencies in answers per factor, measured with the UEQ-S and calculated with the UEQ-S Excel tool [42].

In this fictitious excerpt of responses, it is clear that inconsistencies were found in the answers of three participants. One participant also answered inconsistently in items of both dimensions, which should be critically examined. In addition, two participants ticked the same answer option for all items. One of these two participants only selected the middle option (neutral, 0), which is why it is also recommended to remove answers of this participant.

Another way of assessing how accurate the UX measurements are is reliability, measured using the established statistical method of Cronbach's alpha. The calculations for this are also provided in the Excel tools of the UEQ Family [30], [37], [42]. An example of this is shown in Fig. 11.

The individual values in the Cronbach's alpha column indicate how reliably the questionnaire (in Fig. 11: the UEQ+) measures the respective factors with the items. It can be stated that values > 0.7 are acceptable, while values > 0.8 are good and > 0.9 are very good. As all alpha values in this fictitious example are greater than 0.87, it can be assumed that the measurement accuracy is high. The remaining columns show the correlation of the individual factors. Correlation values of less than 0.9 are desirable here. Values above 0.9 would indicate that individual questions may be too similar to each other or not sufficiently selective.

Scale	Corr(I1,I2)	Corr(I1,I3)	Corr(I2,I3)	Cronbach Alpha
Attractiveness	0,79	0,63	0,66	0,87
Efficiency	0,75	0,81	0,88	0,91
Intuitive Use	0,79	0,85	0,93	0,94
Visual Aesthetics	0,76	0,85	0,85	0,95
Quality of Content	0,46	0,43	0,63	0,83
Trustworthiness of Content	0,58	0,57	0,89	0,90
Social interaction	0,77	0,76	0,87	0,94

Fig. 11. Example of a reliability in answers per factor, measured with the UEQ+ and calculated with the UEQ+ Excel tool [37].

The statistical analyses of the Excel tool can therefore be used to make initial interpretations of the performance of the factors and items, and therefore features of the product under investigation. Nevertheless, in practice, decision-makers are used to having the important information about the company's products and situations summarized in an overall key figure, rather than having to perform various statistical calculations. In addition, as the UEQ+ research basis (see Section III.C) has already shown, different factors are relevant for each product. The question therefore arises as to how the overall UX performance can be summarized, how the importance can be measured and then set in relation to this performance.

B. Benchmark

In addition to the actual statistical measurements relating to how users perceive the UX of products, the first fundamental question in practice is also whether the expectations set in advance have been met. Assuming that the previous statistical measurements produce values and error bars as shown in Fig. 9, the question arises as to whether this is a good or bad result.

It is easy to compare a product version with a previous version to compare expectations and measurements. The UEQ family also provides a separate tool for this purpose free of charge online [30]. However, it becomes more difficult if no previous version is available for comparison, but the question arises as to whether the product has a sufficient UX. It is therefore interesting to compare the extent to which the measured UX of a product matches the measured UX of other products, measured using the same method. Some questionnaires provide benchmarks for this purpose, which contain the UX scores measured with the questionnaire based on a large number of different products. Benchmarks were developed for the UEQ Family in order to provide a greater basis for interpreting the product-specific results.

1. UEQ Benchmark

A benchmark for the UEQ was developed by Schrepp et al. in 2014 and updated in 2017 [3]. Data from a total of 246 products and 9,905 responses were used as a basis for evaluation, including for example business applications, web stores or services, social networks, household appliances and other product types. Due to the high degree of confidence in the actual data, the factor averages were integrated into the data set instead of raw data.

Based on this data set, the UEQ benchmark provides a grouping per factor in the following 5 categories, meaning the evaluated product is

- **Excellent:** among the best 10% of the product base.
- **Good:** 10% of products are better and 75% are worse.
- **Above Average:** 25% of products are better and 50% are worse.
- **Below Average:** 50% of products are better and 25% are worse.
- **Bad:** among the worst 25% of the product base.

With regard to the benchmark dataset basis, 20-30 users already produce stable results. However, general UX expectations change over time. So even if the underlying data set is not continuously updated, new products can still achieve *good* scores [3].

In relation to the results of Fig. 9, this leads to a benchmark orresponding to Fig. 12. The benchmark can be calculated using the UEQ Excel sheet [30].

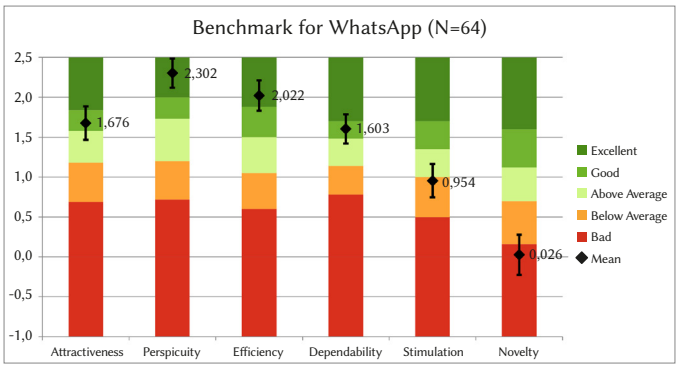

Fig. 12. Product benchmark for WhatsApp (N=64) [25], measured with the UEQ and evaluated with the Excel tool [30].

Compared to the other products of the benchmark dataset, it is clear that the *pragmatic* factors perform in the categories *good* or *excellent*, while the *hedonic* factors can be classified as *below average* and even *bad*. This may be related to the intended use of the product. Even if WhatsApp can be used in a leisure environment, it serves the fulfillment of purposes or completing tasks, such as efficient communication, so that the focus of the product is on the pragmatic properties. For this reason, impact factors (see Section V) should also be taken into account when interpreting the benchmark.

2. UEQ-S Benchmark

As much of the data for the UEQ benchmark comes from practical industrial projects, for confidential reasons the factor means rather than the raw data were included in the creation of the UEQ benchmark. As analyzed by Hinderks et al. [43], it is therefore not possible to synthesize a benchmark exclusively for the 8 items of the UEQ-S on the basis of the data set. However, as it was shown (see Section III.B) that the UEQ values allow a good approximation of the UEQ-S values, the question arose as to whether the UEQ-S benchmark can be calculated on the basis of the UEQ benchmark. For this purpose, the mean values of the factors *Efficiency*, *Perspicuity* and *Dependability* were calculated in a study in 2018 to form the *pragmatic* quality dimension, and the mean values of the *Stimulation* and *Novelty* factors were calculated for the *hedonic* dimension. The overall value was formed on the basis of all UEQ factors, including *Attractiveness*. The following results were thus obtained for the dimensions of the *pragmatic* and *hedonic* quality (PQ and HQ) of the UEQ-S benchmark:

- **Excellent:** PQ greater than 1.73, HQ greater than 1.55, Overall greater than 1.58.
- **Good:** PQ between 1.55 and 1.73, HQ between 1.25 and 1.55, Overall between 1.4 and 1.58.
- **Above Average:** PQ between 1.15 and 1.54, HQ between
- 0.88 and 1.24, Overall between 1.02 and 1.39.
- **Below Average:** PQ between 0.73 and 1.14, HQ between
- 0.57 and 0.87, Overall between 0.68 and 1.01.
- **Bad:** PQ less than 0.73, HQ less than 0.57, Overall less than 0.68.

The benchmark can be calculated again using the corresponding UEQ-S Excel sheet [42]. The results showed that a natural transformation of the UEQ to the UEQ-S is possible in this way. However, over-interpretation is not advisable, as only the factor means were used as a basis for the analyses. More data would be needed to gain a deeper understanding of the relationship between full and short UEQ version.

3. UEQ+ Benchmark

Due to the modular structure of the UEQ+, it is difficult and time-consuming to create a classic benchmark at the level of the individual factors, as only some factors are relevant for certain products. It would require significantly more products and data than when creating the UEQ benchmark. In addition, further UEQ+ factors can be added over time, meaning that a UEQ+ benchmark can never assess all factors with the same quality.

A study from 2023 by Meiners et al. [44] therefore took a different approach and carried out a simple benchmark based on a limited set of product evaluations as quick guidance for UX researchers and practitioners. This benchmark is based on the UX KPI and therefore not on individual factors, but on the overall UX impression of a product. These KPIs also vary depending on the selected factors and products, but the comparison of several products with the same use case is often sufficient in practice to get a first impression of whether one's own product is "good enough" in terms of its UX. Accordingly, for the calculation of the UEQ+ benchmark, which is based on the UEQ+ KPI, various products were evaluated by over 3,200 participants in a total of 26 studies in order to allow an initial understanding of the perceived compared to the expected UX quality of products. Further information can be found in Meiners et al. [44]. The required KPIs can be calculated using the UEQ+ Excel sheet [37]. However, this first UEQ+ benchmark should not be overinterpreted, as this would require more data.

C. Key Performance Indicators

To further effectively evaluate the results of an UX questionnaire, there is a relevant desire for a meaningful key figure, a so-called key performance indicator, as an assessment of the perceived UX of a product. A KPI helps in identifiying important areas for improvement, and the multidimensional nature of the user experience itself already returns multiple scales that provide information about these areas [26]. To now further reduce the complexity of multidimensionality, especially for decision-makers, a UX KPI for the UEQ Family is introduced. In order to enable an assessment of strengths and weaknesses, so-called importance ratings were added to the UEQ from 2019 [26], which respondents use to assess the relevance of the respective factors and associated items for a product. These importance ratings are recorded in a range from -3 to 3. Based on these ratings, the relative importance of each factor can then be calculated for each participant. In order to use a key figure to estimate how strong the correlation is between the perceived quality (the value measured with the UEQ factor) and the perceived importance for the respondents, a dependency is initially assumed. This is understandable, because the more important an item is for a user, the more seriously they will answer it in relation to the product. This can also be seen in Fig. 13.

The *pragmatic* factors of the WhatsApp were rated as more important than the *hedonic* factors. The mean values of the *pragmatic* factors were also rated better than those of the *hedonic* factors. In order to obtain a meaningful overall impression of the product taking these considerations into account, the relative rating of each UEQ factor (measured with the UEQ questions) is multiplied by the relative importance of the same factor (measured with the importance questions), and the individual results are added up and divided by the number of participants. The full explanation can be found in the corresponding study [26].

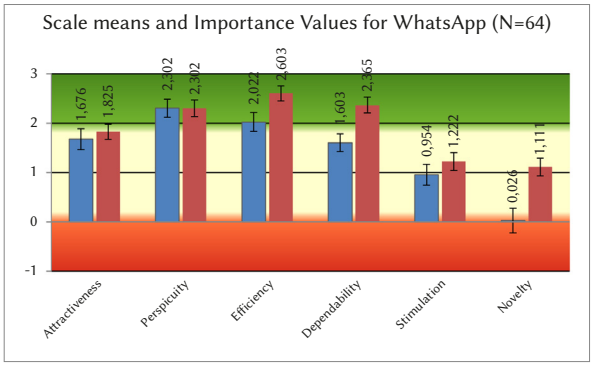

Fig. 13. Product evaluation for WhatsApp (N=64) with factor means (blue) and importance ratings (brown) [25], measured with the UEQ and evaluated with the UEQ Excel tool [30].

The result is again a value in the range from -3 to 3. To simplify the calculation of the KPI, this calculation was also integrated into the Excel tools [30], [37], [42] of the UEQ Family.

However, if we now consider that we want to have a meaningful KPI as an evaluation of the overall UX, then a number alone is not meaningful enough. For example, how do the KPI values 1.1 and 1.4 differ? This raises the question of how an interpretation guideline can be provided for practitioners and researchers. Calculated on the basis of the benchmarks (for futher information see [45]), the KPI value range according to Table I can be determined.

TABLE I. KPI Value Range (Scale From -3 to 3)

UEQ Benchmark	UEQ KPI Min Value	UEQ KPI Max Value	UEQ KPI Mean Value
Excellent	2.018	2.143	2.080
Good	1.375	1.628	1.502
Above Average	1.038	1.290	1.164
Below Average	0.645	0.891	0.768
Bad	-0.286	-0.162	-0.224

It can therefore be summarized that although values from -3 to +3 can theoretically occur, this will most likely not happen in the practical application of the UEQ KPI. The actual range extends from -0.286 to +2.080 and is therefore smaller and more positive. On the basis of this guide, a first impression of the UX of the product or use case under consideration can be gained. However, this key figure should also be treated with caution, as there can be major differences in the importance and evaluation of the actual factors, the performance of the product. For example, a lower rating of the *pragmatic* factors compared to *hedonic* factors, but with a reversed distribution of importance, could also indicate specific weaknesses in the product [26], [45]. It is therefore advisable to take a closer look at the results as part of further analyses.

D. Importance-Performance Analysis

The importance-performance analysis (IPA) [46], [47] is a graphical representation of the relationship between relevance and actual UX assessment of individual factors by participants. The aim of this analysis is to identify specific recommendations for action for the individual factors. It is assumed that a user is satisfied if the perceived importance (brown bar in Fig. 13) has been fulfilled. The fulfillment (performance) is expressed by the factor mean value (blue bar in Fig. 13). This means that it is not the absolute difference between importance and performance which is relevant, but the relative difference between them.

There is no prescribed list of factors for this, which is why this analysis can be carried out with all factors of the UEQ family questionnaires, provided that the importance ratings of the individual factors are also recorded for these factors in the form of a 7-point Likert rating.

According to Hinderks et al. [47], the mean values for each factor are presented in an IPA plot with a total of four quadrants, with each factor being assigned a point. This point is calculated by using the performance value (factor mean value of the perceived UX) for the x-axis and the importance value (mean value of the importance questions) for the y-axis. In this way, the individual quadrants represent concrete recommendations for action, which are described in Table II.

TABLE II. IPA Quadrants [46]–[48]

IPA quadrant	description
Keep Up the Good Work	great strengths and potential competitve advantages both importance and performance equally highly rated no need for action
Possible Overkill	factors rated relatively low, importance is below performance further development of factors not necessary / inefficient
Low Priority	low importance and performance, no action required, balanced
Concentrate Here	most important relatively important, while performance below average [47] highest potential for improvement

An example implementation of this IPA is shown in Fig. 14. Each point in this plot represents a selected factor whose importance and performance (factor mean) was measured with the UEQ+. While the dashed axes represent the original coordinate origin, the solid axes represent the quadrants that are necessary for the IPA interpretation. These are formed by the mean value of all the factors shown and considered [1], [46], [47].

If the importance is higher than the performance, this factor should be improved. In Fig. 14, this applies in particular to the *hedonic* factor *Novelty*. However, no changes are necessary for the *Efficiency* and *Attractiveness* factors. Nevertheless, external factors must also be taken into account when interpreting the results, as can be seen in Fig. 15.

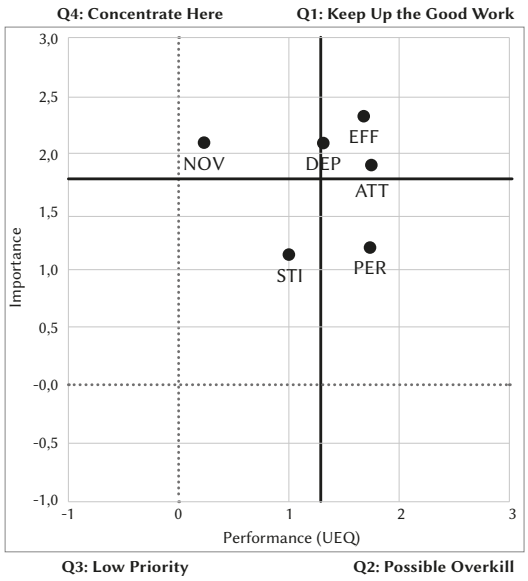

Fig. 14. IPA plot for the product WhatsApp and the factors *Novelty* (NOV), *Stimulation* (STI), *Dependability* (DEP), *Efficiency* (EFF), *Attractiveness* (ATT) and *Perspicuity* (PER), measured with the UEQ+ [1], [46]–[48].

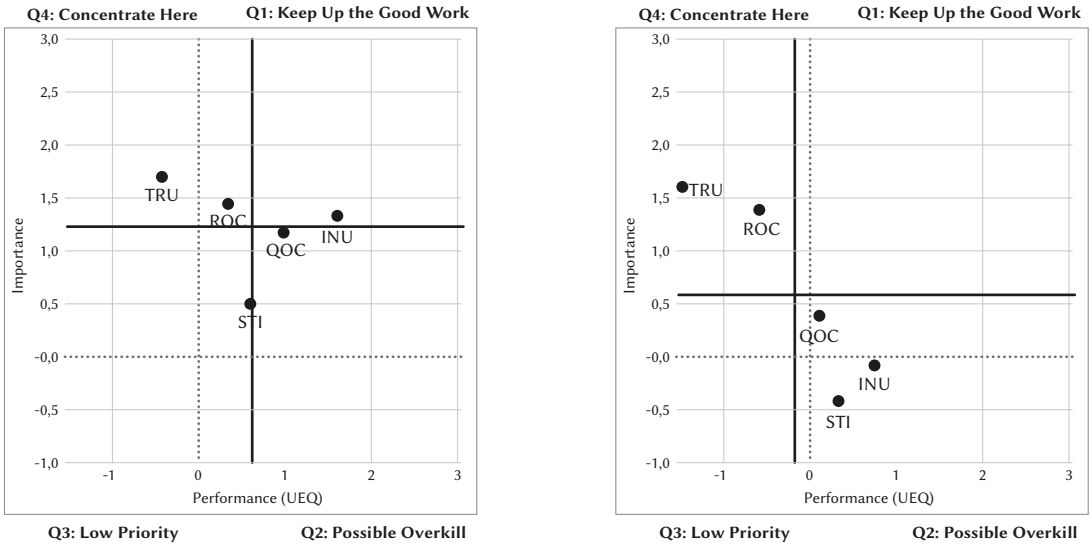

Fig. 15. IPA plot for the product Facebook and the factors *Novelty* (NOV), *Stimulation* (STI), *Dependability* (DEP), *Efficiency* (EFF), *Attractiveness* (ATT) and *Perspicuity* (PER), measured with the UEQ+ [1], [46]–[48].

Fig. 15 compares two IPAs, where variant (a) was created on the basis of users who use the fictitious product every day and variant (b) on the basis of users who do not use it every day. It is clear that the frequency of use can therefore influence the interpretation of the statistical analyses, benchmark, KPI and IPA, and thus the recommendations for action. For this reason, relevant impact factors that are necessary for understanding the UX measurement with the UEQ family are presented in the following Section.

V. Impact Factors

Over the years, various studies have been carried out on impact factors on the perceived user experience. A number of aspects were found to have an influence.

A. User-Related Influences

Starting with the factor **gender**, it cannot be ruled out that, depending on the product type, this may have an influence on the perceived user experience measured with the UEQ, even if it can be small [49], [50]. These results were also confirmed in comparison with other questionnaires. Therefore, if a specific target group is defined on the basis of gender, it should be ensured in advance that the expectations can be met accordingly.

The situation is similar with external influencing factors such as **duration of use**, **frequency of use** and **knowledge** of the product [50], [51]. It can be assumed that users who use their product frequently typically know it better and adapt their usage behavior accordingly in order to avoid typical UX problems with the product. Accordingly, the perceived user experience can also change, which could be confirmed on the basis of various products. The influences of frequency of use and knowledge have been proven, but these also vary depending on the specific product. Accordingly, it is also advisable to select the right factors for each product or product category in advance in order to make the user experience more suitable (see Section V).

The **cultural influence** should not be ignored either. As various studies have shown, there are differences in the perceived UX, which can be explained by different cultures. This should be distinguished from the perceived importance of UX factors, which were shown in these studies to be dependent on the product or product category (see Section V), as there were clear similarities between the cultures investigated. However, differences in perception within the German vs. the Indonesian culture were found, even if they were only minor. The greatest differences were found in the perception of *hedonic* factors such as *Identity*, *Novelty* and *Stimulation*. This was demonstrated with the help of the proportion of variance, which was proven to be generally higher for *hedonic* than *pragmatic* UX factors. These *hedonic* factors were rated relatively low by the German participants compared to the *pragmatic* factors. Nevertheless, also the extent of the influence of culture compared to other interindividual differences between persons was examined here. It was found that the influence of culture is relatively low compared to the impact of differences on an individual level. On the one hand, this means that culture is an impact factor that should not be ignored when measuring UX. On the other hand, however, it should not be overestimated, as personal preferences predominate. Studies on the influence of other cultures should therefore be carried out before further conclusions are drawn here [52]–[54].

Finally, there are influences that relate to the structure of the questionnaires depending on the person who is filling it out. The circumstances are different in contrast to the personal influences, for example with regard to the **item polarity** of semantic differentials. While the UEQ has a mixed item polarity, in which half of the positive items (e.g. entertaining, easy to use) are randomized to the right and the other half to the left, the UEQ-S and UEQ+ have a one-sided (right-sided) item polarity. The basic idea of the study by Schrepp et al. in 2023 [25] was to simplify the processing of the UEQ for the participants by using a one-sided variant. However, studies have shown, also in comparison with other questionnaires, that a change in polarity of the UEQ to a one-sided variant has more disadvantages than advantages. Even if the number of inconsistencies, clicks and the processing time is slightly lower with a one-sided variant, these effects were proven to be very small practically irrelevant [25]. Furthermore, a response tendency is introduced in which participants tend to place items on the same side if they are uncertain about individual items (e.g., on the right if they have a positive perception of the product). Due to this conditional influence of item polarity, a change in the UEQ should therefore be avoided [25].

B. Product-Related Influences

In addition to the influencing factors that relate to the users and their environment, influencing factors that relate to the products under consideration are also clearly present. It has been proven that the user experience is evaluated differently depending on the **product** or **product category**. UX is generally difficult to measure not only because different users subjectively evaluate the same product differently, but also because the same user evaluates different products differently. This is due to the association of products with certain characteristics. For example, while the *Quality of Content* is not applicable for word processing tools, it is most relevant for news portals. The factor of *Stimulation* is also rated as less relevant for online banking platforms, in contrast to social networks (see Fig. 16). This must be taken into account both when measuring and evaluating the perceived UX using the UEQ Family.

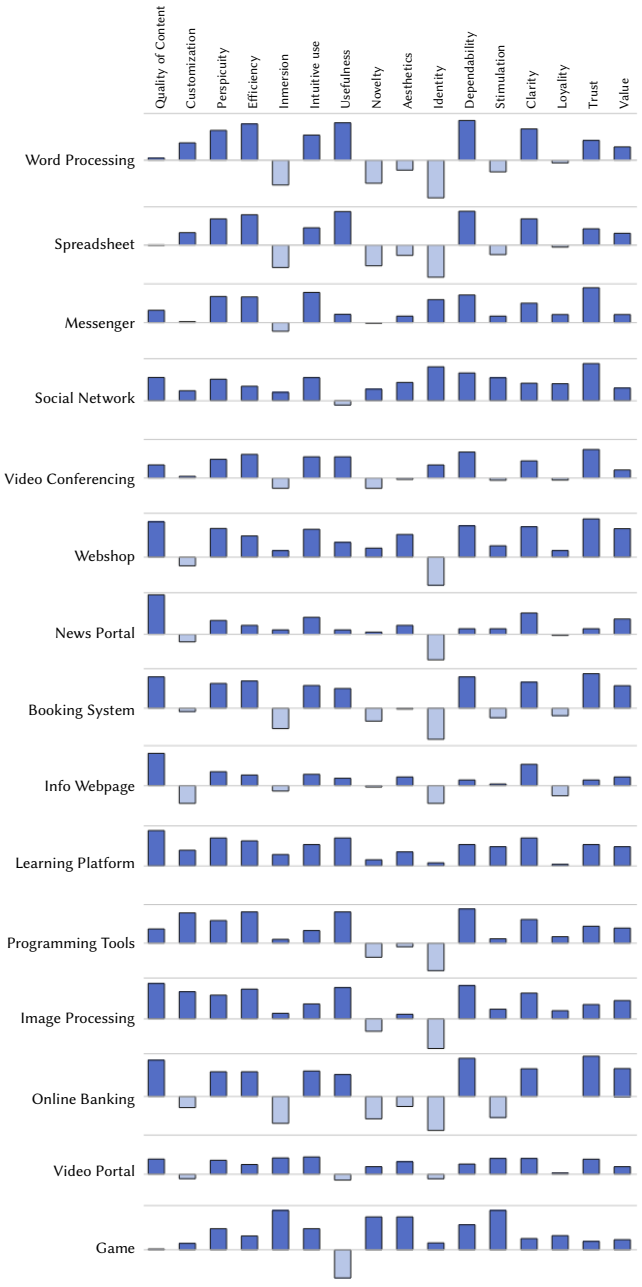

Fig. 16. Means of the importance ratings for the UX quality aspects per product category [32]. Scale ranges from -3 to +3.

These importance ratings can also be transferred to specific products. Studies have shown high to almost perfect correlations between the importance ratings of specific products and their associated product categories. These include, for example, Microsoft Word in the Text Processing category (see Fig. 17), WhatsApp in the Messengers category, or YouTube and Netflix in the Video Portals category [55].

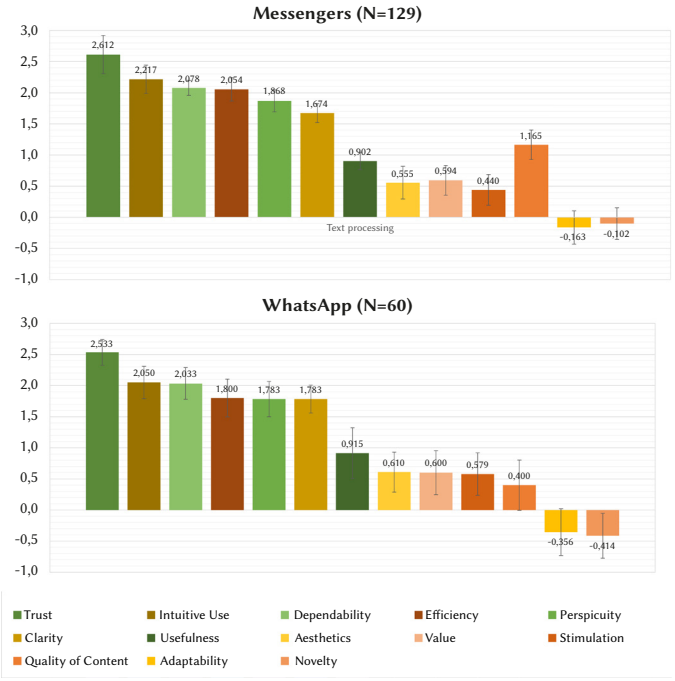

Fig. 17. Importance ratings for the product WhatsApp and the product type Word Messengers. Scale ranges from -3 to +3.

A similar analysis was carried out for the product category of collaboration tools. Here, for example, the UX factors *Trust*, *Perspicuity* and *Efficiency* were rated as the most important overall. The individual differentiations of the products themselves, e.g., Microsoft Teams or Discord, did not impact the rating of the UX aspects as much [56].

The question that now arises is how these impact factors and interpretation notes can be taken into account in practice.

VI. Conclusion and Future Work

Measuring user experience is useful in order to record ongoing improvements to products, services and systems that may result from continuous bug fixes or new releases. Furthermore, it is essential for the product success to determine whether the UX is sufficient, where there are areas for improvement and also whether the investments in the UX are worthwhile in terms of return on investment [57]. UX heuristics already form a suitable basis and guideline for this [18].

Nevertheless, it can be difficult to decide what exactly should be measured. On the one hand, users have different subjective and product-dependent opinions. On the other hand, it is necessary to determine what exactly is meant by UX and to select the appropriate factors for the use case and the research questions from a large number of possible factors. For this reason, standardized questionnaires such as the UEQ, UEQ-S and UEQ+ help to determine what exactly a user thinks about a product and how the generally perceived UX can be assessed.

With the help of semantic differentials, the UEQ forms a good basis for a broad range of use cases and at the same time provides

meaningful possibilities for interpretation. The UEQ is currently available in more than 30 languages [33] and also in a simplified language form, e.g., for children [28].

However, it is still possible that the UEQ takes up too much time for the specific target group. In this case, the UEQ-S can save time, but provides a less detailed picture. In cases where a comprehensive impression for a more specific research question is needed, the usage of the UEQ+ is recommended. Modular questionnaires like the UEQ+ are suitable if either a complete picture of a product needs to be formed, or also only specific UX factors need to be considered. With the help of the UEQ+, it is possible to avoid using other questionnaires in addition to the UEQ in order to capture more than the predefined factors. It also makes it possible to select specific factors depending on the product. Also the UEQ+ is currently available in more than 20 languages, so that it offers potential for versatile use.

However, both when using modular and holistic questionnaires from the UEQ Family, is recommended to consider in advance which product or product category is to be evaluated or given priority. Existing studies provide indications of the importance of certain UX factors for specific products and categories. These are advisable to be used when selecting the appropriate factors. Even if individual products have possibly not yet been considered in explicit studies, it has been confirmed that the UX factors relevant to certain product categories can be used to draw conclusions about individual products [32], [55] and are among the good practices of creating a common UX vision and shared understanding [58]. These findings can also be integrated into internal company processes. For example, the selection of suitable UX factors for each product can be integrated into agile methods such as UX Poker [59]. This method takes place early on in the process and facilitates a shared understanding of UX within the team by assessing the influence of user stories based on selected UX factors. The results of the UX evaluation can thus contribute directly to fulfilling the requirements of UX management to create a positive UX [60].

Once the appropriate factors have been selected and the user data collected, it is necessary to statistically analyze the results. Section IV and the Excel tools provided by the UEQ family [30], [37], [42] offer guidance on how to interpret the results appropriately. This means that no previous experience with questionnaires is necessary to facilitate the use and implementation of the UEQ Family. In addition, however, this article presents common interpretation methods (e.g. KPI and IPA) as well as concrete interpretation aids (e.g. benchmark) for the results.

In addition, the UEQ family can both supplement questionnaires already in use (e.g. on usability aspects or KPIs) and replace them (e.g. by providing a more holistic picture through the addition of a hedonic perspective). Existing results often do not have to be discarded, but can be derived using the UEQ (e.g. SUS or UMUX-LITE, see Section III.D). The UEQ family can therefore be used in a variety of ways depending on the research question and company context.

References

[1] A. Hinderks, D. Winter, M. Schrepp, J. Thomaschewski, "Applicability of user experience and usability questionnaires," *Journal of Universal Computer Science*, no. 25, pp. 1717–1735, 2020.

[2] E. L.-C. Law, P. van Schaik, "To measure or not to measure ux: An interview study," in *International Workshop on the Interplay between User Experience Evaluation and System Development (I-UxSED)*, 2012.

[3] M. Schrepp, J. Thomaschewski, A. Hinderks, "Construction of a benchmark for the user experience questionnaire (ueq)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 4, pp. 40–44, 2017, doi: 10.9781/ijimai.2017.445.

[4] M. Schrepp, A. Hinderks, J. Thomaschewski, "Applying the user experience questionnaire (ueq) in different evaluation scenarios," in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, Cham, 2014, pp. 383–392, Springer International Publishing. isbn: 978-3-319- 07668-3.

[5] M. Schrepp, J. Thomaschewski, "UEQ+ Handbook." https://ueqplus.ueq-research.org/Material/ UEQ+_Handbook_V6.pdf, 2023. [Online; accessed 24-April-2024].

[6] M. Schrepp, *User Experience Questionnaires: How to use questionnaires to measure the user experience of your products?* Self-published, 2021. isbn: 979-8736459766.

[7] M. Schrepp, A. Hinderks, J. Thomaschewski, "Design and evaluation of a short version of the user experience questionnaire (ueq-s)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, p. 103, 2017, doi: 10.9781/ijimai.2017.09.001.

[8] B. Laugwitz, T. Held, M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, vol. 5298 of *Lecture Notes in Computer Science*, A. Holzinger Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76, doi: 10.1007/978-3-540-89350-9_6.

[9] M. Schrepp, J. Thomaschewski, "Design and validation of a framework for the creation of user experience questionnaires," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 7, 2019, doi: 10.9781/ijimai.2019.06.006.

[10] DIN EN ISO 9241-11, "Ergonomics of human-system interaction — part 11: Usability: Definitions and concepts," 2018. doi: 10.31030/2757945.

[11] ISO9241-210, "Ergonomics of human-system interaction - part 210: Human-centred design for interactive systems," 2020. doi: 10.31030/3104744.

[12] M. Hassenzahl, "The thing and i: Understanding the relationship between user and product," in *Funology*, M. A. Blythe Ed., Boston [etc.]: Kluwer Academic Publishers, 2003, pp. 31–42. isbn: 978-1402029660.

[13] M. Schrepp, J. Kollmorgen, J. Thomaschewski, "A comparison of sus, umux-lite, and ueq-s," *Journal of User Experience*, vol. 18, p. 86–104, jun 2023.

[14] M. P. Cota, J. Thomaschewski, M. Schrepp, R. Gonçalves, "Efficient measurement of the user experience. a portuguese version," *Procedia Computer Science*, vol. 27, pp. 491–498, 2014, doi: 10.1016/j.procs.2014.02.053.

[15] V. Roto, E. L.-C. Law, A. Vermeeren, J. Hoonhout, "User experience white paper: Bringing clarity to the concept of user experience," Schloss Dagstuhl - Leibniz- Zentrum für Informatik, 2011.

[16] E. M. Schön, J. Hellmers, J. Thomaschewski, "Usability evaluation methods for special interest internet information services," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. 6, pp. 26–32, 2014, doi: 10.9781/ijimai.2014.263.

[17] J. Nielsen, R. Molich, "Heuristic evaluation of user interfaces," in *CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1990, pp. 249–256. isbn: 0201509326.

[18] F. Bader, E.-M. Schön, J. Thomaschewski, "Heuristics considering ux and quality criteria for heuristics," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2017, doi: 10.9781/ijimai.2017.05.001.

[19] 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*, vol. 2, no. 1, p. 39, 2013, doi: 10.9781/ijimai.2013.215.

[20] M. Schrepp, "A comparison of ux questionnaires - what is their underlying concept of user experience?." Mensch und Computer 2020 - Workshopband, Bonn, 2020. doi: 10.18420/muc2020-ws105-236.

[21] J. Brooke, "Sus: A quick and dirty usability scale," in *Usability Evaluation In Industry*, CRC Press, 1996, p. 6. isbn: 9780429157011.

[22] M. Schrepp, "UEQ Handbook." https://www. ueq-online.org/Material/Handbook.pdf, 2023. [Online; accessed 24-April-2024].

[23] M. Hassenzahl, M. Burmester, F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Mensch & Computer 2003*, Stuttgart [u.a.], 2003, pp. 187–196, Teubner. isbn: 978-3519004417.

[24] M. Minge, L. Riedel, "mecue - ein modularer fragebogen zur erfassung des nutzungserlebens," in *Mensch & Computer 2013: Interaktive Vielfalt*, München, 2013, pp. 89–98, Oldenbourg Verlag. doi:

10.1524/9783486781229.89.

[25] M. Schrepp, J. Kollmorgen, J. Thomaschewski, "Impact of item polarity on the scales of the user experience questionnaire (ueq)," in *International Conference on Web Information Systems and Technologies*, 2023, pp. 15–25. doi: 10.5220/0012159900003584.

[26] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, J. Thomaschewski, "Developing a ux kpi based on the user experience questionnaire," *Computer Standards & Interfaces*, vol. 65, pp. 38–44, 2019, doi: https://doi.org/10.1016/j.csi.2019.01.007.

[27] A. Hinderks, M. Schrepp, J. Thomaschewski, "UEQ and UEQ-S Website." https://www.ueq-online.org/, 2024. [Online; accessed 24-April-2024].

[28] M. Schrepp, M. P. Cota, R. M. Gonçalves, A. Hinderks, J. Thomaschewski, "Adaption of user experience questionnaires for different user groups," *Universal Access in the Information Society*, vol. 16, pp. 629 – 640, 2016, doi: 10.1007/s10209-016-0485-9.

[29] M. Hernández-Campos, J. Thomaschewski, Y. C. Law, "Results of a study to improve the spanish version of the user experience questionnaire (ueq)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 4, pp. 202–207, 2023, doi: 10.9781/ijimai.2022.11.003.

[30] A. Hinderks, M. Schrepp, J. Thomaschewski, "UEQ Data Analysis Tool." https://www.ueq-online.org/ Material/Data_Analysis_Tools.zip, 2024. [Online; accessed 24-April-2024].

[31] A. Hinderks, M. Schrepp, M. Rauschenberger, J. Thomaschewski, "Reconstruction and validation of the ux factor trust for the user experience questionnaire plus (ueq+)," in *Proceedings of the 19th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST*, 2023, pp. 319–329, INSTICC, SciTePress. doi: 10.5220/0012186700003584.

[32] M. Schrepp, J. Kollmorgen, A. Meiners, A. Hinderks, D. Winter, H. Santoso, J. Thomaschewski, "On the importance of ux quality aspects for different product categories," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 2, pp. 232–246, 2023, doi: 10.9781/ijimai.2023.03.001.

[33] M. Schrepp, J. Thomaschewski, "UEQ+ Website." https://ueqplus.ueq-research.org/, 2024. [Online; accessed 24-April-2024].

[34] A. M. Klein, A. Hinderks, M. Schrepp, J. Thomaschewski, "Construction of ueq+ scales for voice quality: measuring user experience quality of voice interaction," in *Mensch und Computer 2020 - Tagungsband*, New York: ACM, 2020, p. 1–5, doi: 10.1145/3404983.3410003.

[35] M. Lindemann, H. Weber, "Ux in zahlen: Der user experience questionnaire for medical devices," in *Usability Professionals 23*, Gesellschaft für Informatik e.V., 2023, doi: 10.18420/muc2023-up-455.

[36] E. Mortazavi, P. Doyon-Poulin, D. Imbeau, J.-M. Robert, "Development and validation of four social scales for the ux evaluation of interactive products," *International Journal of Human-Computer Interaction*, pp. 1–14, 2023, doi: 10.1080/10447318.2023.2258026.

[37] M. Schrepp, J. Thomaschewski, "UEQ+ Data Analysis Tool." https://ueqplus.ueq-research. org/Material/UEQ_Plus_Data_Analysis_Tool.xlsx, 2024. [Online; accessed 24-April-2024].

[38] A. Hinderks, M. Schrepp, J. Thomaschewski, "Vergleich von ux fragebögen." doi: 10.18420/muc2018-mci-0363.

[39] J. R. Lewis, B. S. Utesch, D. E. Maher, "Umux-lite: when there's no time for the sus," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, New York, NY, USA, 2013, p. 2099–2102, Association for Computing Machinery. doi: 10.1145/2470654.2481287.

[40] U. Lah, J. R. Lewis, B. Sumak, "Perceived usability and the modified technology acceptance model," *International Journal of Human–Computer Interaction*, vol. 36, pp. 1216 – 1230, 2020, doi: 10.1080/10447318.2020.1727262.

[41] A.-L. Meiners, A. Hinderks, J. Thomaschewski, "Korrelationen zwischen ux-fragebögen." Mensch und Computer 2020 - Workshopband, Bonn, 2020. doi: 10.18420/muc2020-ws105-375.

[42] A. Hinderks, M. Schrepp, J. Thomaschewski, "UEQ-S Data Analysis Tool." https://www.ueq-online.org/ Material/Short_UEQ_Data_Analysis_Tool.xlsx, 2024. [Online; accessed 24-April-2024].

[43] A. Hinderks, M. Schrepp, J. Thomaschewski, "A benchmark for the short version of the user experience questionnaire," in *Proceedings of the 14th International Conference on Web Information Systems and Technologies (WEBIST 2018)*, 2018, pp. 373–377, SCITEPRESS - Science and Technology Publications. doi: 10.5220/0007188303730377.

[44] A.-L. Meiners, M. Schrepp, A. Hinderks, J. Thomaschewski, "A benchmark for the ueq+ framework: Construction of a simple tool to quickly interpret ueq+ kpis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, pp. 1–8, 2023, doi: 10.9781/ijimai.2023.05.003.

[45] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, J. Thomaschewski, "Ueq kpi value range based on the ueq benchmark." doi: 10.13140/RG.2.2.34239.76967.

[46] A. Hinderks, A.-L. Meiners, F. Mayo, J. Thomaschewski, "Interpreting the results from the user experience questionnaire (ueq) using importance-performance analysis (ipa)," in *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, WEBIST 2019, Setubal, PRT, 2019, p. 388–395, SCITEPRESS - Science and Technology Publications, Lda. doi: 10.5220/0008366503880395.

[47] A. Hinderks, F. J. Domínguez-Mayo, A.-L. Meiners, J. Thomaschewski, "Applying importance- performance analysis (ipa) to interpret the results of the user experience questionnaire (ueq)," *Journal of Web Engineering*, 2020, doi: 10.13052/jwe1540- 9589.1926.

[48] A. Hinderks, "A lifecycle for user experience management in agile development." PhD Thesis, Universidad de Sevilla, Sevilla, 2021.

[49] M. Schrepp, J. Thomaschewski, K. Aufderhaar, "Do women and men perceive user experience differently?," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 6, pp. 63–67, 2019, doi: 10.9781/ijimai.2019.03.005.

[50] J. Kollmorgen, M. Schrepp, J. Thomaschewski, "Influence of demographic variables and usage behaviour on the perceived user experience," in *Web Information Systems and Technologies*, Cham, 2023, pp. 186–208, Springer Nature Switzerland. isbn: 978-3-031-43088-6.

[51] J. Kollmorgen, M. Schrepp, J. Thomaschewski, "Impact of usage behaviour on the user experience of netflix, microsoft powerpoint, bigbluebutton and zoom," in *International Conference on Web Information Systems and Technologies*, 2022, pp. 297–406. doi: 10.5220/0011380100003318.

[52] H. Santoso, M. Schrepp, A. Hinderks, J. Thomaschewski, "Cultural differences in the perception of user experience." Mensch und Computer 2017 - Tagungsband, Regensburg, 2017. doi: 10.18420/muc2017-mci-0272.

[53] M. Schrepp, H. Santoso, "Has culture an impact on the importance of ux aspects?," in *Mensch und Computer 2018 - Workshopband*, Bonn: Gesellschaft für Informatik e.V., 2018, doi: 10.18420/muc2018-ws04-0473.

[54] H. B. Santoso, M. Schrepp, "The impact of culture and product on the subjective importance of user experience aspects," *Heliyon*, vol. 5, no. 9, p. e02434, 2019, doi: https://doi.org/10.1016/j.heliyon.2019.e02434.

[55] A.-L. Meiners, J. Kollmorgen, M. Schrepp, J. Thomaschewski, "Which ux aspects are important for a software product? importance ratings of ux aspects for software products for measurement with the ueq+.," in *Proceedings of Mensch Und Computer 2021*, MuC '21, New York, NY, USA, 2021, p. 136–139, Association for Computing Machinery. doi: 10.1145/3473856.3473997.

[56] A.-L. Meiners, A. Hinderks, J. Thomaschewski, "Trust, perspicuity, efficiency: Important ux aspects to consider for the successful adoption of collaboration tools in organisations," in *Computer-Human Interaction Research and Applications*, Cham, 2023, pp. 143–162, Springer Nature Switzerland. isbn: 978-3-031-49425-3.

[57] M. Schrepp, "Measuring user experience with modular questionnaires," in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2021, pp. 1–6. doi: 10.1109/ICACSIS53237.2021.9631321.

[58] D. Winter, C. Hausmann, A. Hinderks, J. Thomaschewski, "Development of a shared ux vision based on ux factors ascertained through attribution," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 2, pp. 247–254, 2023, doi: 10.9781/ijimai.2023.04.001.

[59] A. Hinderks, D. Winter, F. J. D. Mayo, M. J. Escalona, J. Thomaschewski, "Ux poker: Estimating the influence of user stories on user experience in early stage of agile development," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 7, pp. 97–104, 2022, doi: 10.9781/ijimai.2022.11.007.

[60] A. Hinderks, F. J. D. Mayo, M. J. Escalona, J. Thomaschewski, "Requirements for user experience management - a tertiary study," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 6, pp. 160-167, 2024, doi: 10.9781/ijimai.2024.01.004.

Jessica Kollmorgen

Jessica Kollmorgen received a B.Sc. degree in Business Informatics, with parallel employment in backend software development, and a Master's degree in Media Informatics with a specialization in Mobile Computing and Security at the University of Applied Sciences Emden/Leer, Germany. Currently, she is a part-time research associate in the area of Agile User Experience. She has been an active member of the research group 'Agile Software Development and User Experience' since 2021. Her research interests lay on user experience factors, UX measurement and statistics as well as Agile UX.

Andreas Hinderks

Andreas Hinderks holds a Ph.D. in Computer Science by University of Seville (Spain). He became a Full Professor at the University of Applied Sciences and Arts Hannover (Germany) in March 2024. He has worked in various management roles as a Business Analyst and a programmer from 2001 to 2016. His focus lays on developing user-friendly business software. Currently, he is a freelancing Product Owner, Business Analyst and Senior UX Architect. He is involved in research activities dealing with UX questionnaires, measuring user experience and user experience management since 2011.

Jörg Thomaschewski

Jörg Thomaschewski received a Ph.D. in physics from the University of Bremen (Germany) in 1996. He became a Full Professor at the University of Applied Sciences Emden/Leer (Germany) in September 2000. His teaching and research focus is on Human-Computer Interaction, UX Management, Agile Software Development, and Requirements Engineering. In 2012 he founded the research group 'Agile Software Development and User Experience'. Dr. Thomaschewski has extensive experience in user experience training, UX questionnaires, agile methods, IT analysis, and consulting.