User Generated Inverted Item Evaluation (UGIIE)

Denys J.C. Matthies[1], Jagankumar Kothandan[2], Wenhan Sun[1,3], and Anke Dittmar[2]

[1] Technical University of Applied Sciences Lübeck, Germany
[2] University of Rostock, Germany
[3] East China University of Science and Technology, Shanghai, China

Abstract. Measuring short-term User Experience (UX) of interactive systems is often assessed by standardized questionnaire-based methods. As standardized methods are broadly applicable, they deploy a great variety of questions. Depending on the target system, certain questions may appear irrelevant and annoying, while others may even bias the participant. This paper proposes an alternative UX evaluation method, the User Generated Inverted Item Evaluation (UGIIE). It utilizes a think-aloud technique to elicit only relevant items for a system-tailored questionnaire and employs an item inversion to increase critical reflection. Utilizing four user studies, UGIIE is benchmarked through a comparative study with two common UX methods, meCUE and UEQ. We found UGIIE to decrease users' frustration when filling out the questionnaire. UGIIE also indicates to provide a higher consistency in feedback when comparing the participants' quantitative rating and qualitative comments. UGIIE's drawback is the rather time-consuming preparation. Our study results suggest a preparation with a minimum of five participants to discover at least 80% of all relevant items for a UGIIE questionnaire. However, when testing much larger numbers of participants, UGIIE can be executed without the presence of an experimenter and thus becomes increasingly efficient.

Keywords: User Experience (UX) · Evaluation Method / Technique

1 Introduction

Research and development in the User Experience (UX) field is constantly growing, receiving high interest from industries and academia. UX is generally described as a result of an interaction between the user, the system and the context [1]. According to the ISO 9241-210, UX is defined as a person's *"perceptions and responses that result from the use or anticipated use of a product, system or service"* [2]. This formal definition is supplemented by other interpretations where the user experience *"also explores how a person feels about using a product, i.e., the experiential, meaningful and valuable aspects of the product use"* [3].

These definitions therefore indicate that measuring a realistic UX impression may be challenging. This is because the person's *feelings* and *perceptions* are highly subjective, manifold, individual, and thus ambiguous. Consequently, researchers and practitioners have developed a vast collection of standardized evaluation methods. To an extent, these methods measure a person's perception of the system before, during, and after the interaction. These UX evaluation methods are mostly questionnaire-based and aim to be broadly applicable. However, based on practitioners experiences, the results can occasionally be ambiguous and inaccurate. Other issues also arise where UX questionnaires are perceived as being inappropriate, as a number of questions fail to apply to the current tested system.

This paper proposes a new evaluation method, the *User Generated Inverted Item Evaluation (UGIIE)*. This method focuses on measuring short term UX by targeting a single behavioral episode having a defined beginning and end. The quality of the proposed method has been benchmarked by a comparative evaluation approach, in which the participants completed four diverse tasks. Each task involved an interactive product (1. flying a drone, 2. handling an unknown smartphone application, 3. setting up a mobile projector, and 4. using Instagram on a smartphone). The participant's UX was measured by: *User Experience Questionnaire (UEQ)*, *Modular Evaluation of Key Components of UX (meCUE)*, and *User Generated Inverted Item Evaluation (UGIIE)*.

Although UGIIE is also a questionnaire-based evaluation method, its core idea differentiates it from existing methods. UGIIE is based on the concept of *Reverse Brainstorming* [4], which is a creative technique that attempts to locate constructive ideas and solutions by simply inverting the goal. This deception releases the human brain from thought processes reliant on old structures, enabling the participant to reflect on the topic or system from a different angle. This allows a more creative problem-solving [5] and explains why UGIIE tends to achieve a slightly more critical reflection, resulting in greater consistency in the feedback of the actual UX during the study and final quantitative rating. Based on our results, it is suggested to run UGIIE with at least five participants individually to extract the required parameters for the inverted item questionnaire (*see Figure 10*). UGIIE seems to only apply relevant items, which also decreases the users' frustration when filling out the questionnaire. Considering we utilized a combination of qualitative and quantitative analysis, results are still extracted in a reasonably short amount of time.

2 Background

2.1 Understanding User Experience

The design of products and services increasingly focuses on user enjoyment, while simultaneously supporting fundamental human needs and values [6, 7]. Meanwhile, User Experience (UX) exists as a core aspect of product development [8]. UX is generally understood as inherently dynamic, given the ever-changing internal and emotional state of a person and difference in the circumstances

during and after an interaction with a product [9, 10]. While it is relevant to evaluate short-term experience [11],investigating the temporal change of UX [10, 12] is also essential. When referring to the long-term use of interactive systems, products, and services, Kujala et al. [13] state the importance of the following UX attributes: attractiveness of the system, ease of use, utility, and degree of usage.

In HCI, understanding UX is regarded as an important issue. Several techniques, such as interviews, observations, surveys, story-telling, and diaries among others [14] have been explored. Numerous peripheral factors, such as peer groups, used products, and the environment substantially influence the UX that the interaction evokes [15]. The following aspects directly influence the experience evoking directly from user-product interaction: individual values, emotions, expectations, and prior experiences among similar products [16, 17]. The "best-practise" to measure such attitudinal data is to conduct it either on a small scale, such as in the lab, or on a large scale by using surveys [18, 19]. Within this century, the HCI field has substituted usability concerns with the UX. A methodological shift from a quantitative to a qualitative approach occurred, as noted by Bargas et al. [20].

2.2 Relationship of User Experience and Usability

User Experience (UX) and usability are fundamental for a successful product and service delivery. According to the ISO 9241-11, usability is defined as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [21]. Long time the term UX remained vague and was widely used as a synonym for other terms [22]. This changed when the concept of usability was defined by ergonomics research in the early 1980's. The concept of usability, however, has gradually evolved into a definition that captures the quality of use [23]. Hertzum [24] describes six different perspectives on usability: universal usability, situational usability, perceived usability, hedonic usability, organizational usability, and cultural usability.

In contrast, UX focuses on the individual experience, rather than effectively and efficiently achieving a goal within the context of product use [9, 20, 23]. Hassenzahl [25] and Bevan et al. [23], however, point out that usability and UX are both underpinned by an element of satisfaction. Therefore, this clarification can be added to the ISO 9241-11. Nevertheless, differences exist between UX and usability. UX focuses on lived experiences [26], whereas usability focuses on evaluating task performance. Also, UX is highly subjective [10]. Usability, however, is objectively measurable using typical measures, such as the task completion time, number of clicks, error rate, etc. Another way to measure usability is considering the user's *"satisfaction"*, which is also a core aspect of UX evaluations. As the UX addresses a range of other subjective qualities, usability may also be a subset of the UX [22].

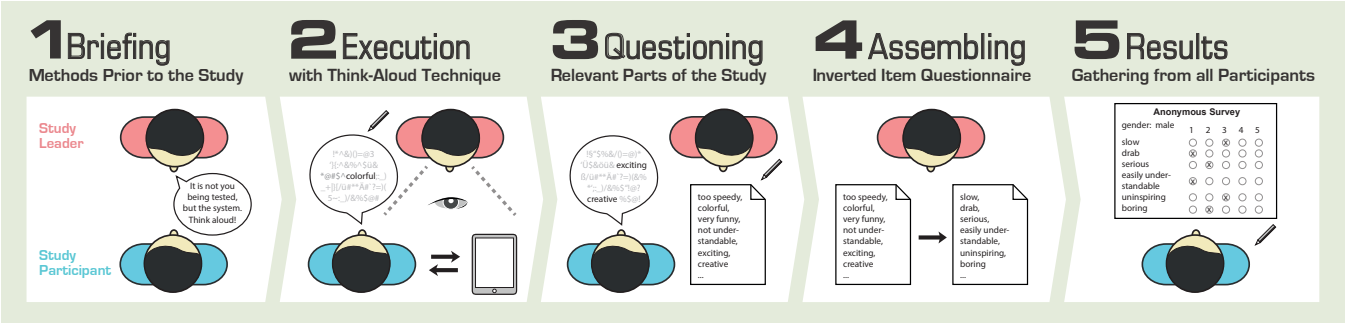

Fig. 1. The UGIIE technique features five phases. In a short (1) Briefing the participant is explained important considerations, such as to think aloud. At the (2) Execution phase of the study, the user is interacting with the target system. A (3) Questioning post study helps to elicit additional attributes describing the subject's experience. (4) Assembling the inverted item questionnaire is a crucial step. Once the questionnaire is compiled, the (5) Results are gathered from all participants.

3 User Generated Inverted Item Evaluation (UGIIE)

This method is based on initial qualitative user feedback [27], which is then transformed into quantitative measurable results. Using clear user instructions and classic methods, this technique has potential to overcome common drawbacks of standardized quantitative methods. It particularly circumvents the problem of user's failing to independently undertake a critical reflection by assessing a semi-negative questionnaire that is compiled from the user's gathered feedback. This methodology appropriates the idea of a creativity technique, the *"Brainstorm Paradox / Reverse Brainstorming"* [4], which attempts to discover constructive ideas and solutions through goal inversion. The human brain is released from relying on old structures, enabling the participant to view the topic from a different perspective. This has been discovered and utilized for creative problem solving as early as 1974 [5]. To perform a successful evaluation, a study leader and a suggested minimum of five participants are required. This minimum enables the finding of at least ∼80% of all relevant items (*see Figure 10*). However, we suggest a number of eight participants. In fact, the sample size between eight and twelve will also yield a larger finding of the majority of all usability problems [28, 29]. To improve post-processing and subsequent analysis, it is suggested to include audio and video recordings, while having an additional observer taking notes. Once the questionnaire is compiled, UGIIE can be executed without the presence of the experimenter.

3.1 Classification

From the later introduced criteria *(see section: Related Work)* based on Vermeeren's taxonomy [3], UGIIE could be classified as follows:

Origin: Similar to most UX assessment methods, the UGIIE framework was developed in a scholarly context.

Type of Gathered Information: Although UGIIE relies on qualitative feedback [27], it transforms it into quantitative feedback, which can be analyzed by statistical means.

Type of Applications/Designs: UGIIE is applicable to a large variety of applications, such as testable multimedia information systems (e.g., smartphone interfaces, games, interactive installations etc.). While other application fields (e.g., on-site with users) may be possible, they have yet to be tested.

Data Sources: The proposed evaluation method is based on the *"think aloud method"* [30] and can be performed individually or in groups. A specific user group is not necessary. However, an *"expert"* would need to extract the attributes and compile the questionnaire.

Location: Any location, such as a lab environment or the field is possible that allows the study leader to track the user's experiences using the *"think aloud method"* [30]. Remote-environments, namely when the participant is at a different location and using a web-based service, may create complications, but is not impossible using a technical workaround.

Period of Experience: UGIIE operates most effectively at a single behavioural episode with a defined beginning and end (e.g., task or period in which the user explores some specific feature). It also works for a typical test session (e.g., one hour of performing a task).

Development Phases: The proposed method is sufficient for interactive products, such as fully functional products and functional prototypes.

Technical Requirements: UGIIE does not require any specialized equipment. However, it requires a *"trained UX examiner"*, to extract the items and to prepare the questionnaire. Conducting it remotely requires technical applications.

The method itself consists of five phases; briefing, execution, questioning, assembling the attribute list, and gathering results. The work-flow is explained in the subsequent sections.

3.2 I. Briefing

Prior to executing the user study, a short briefing phase is required. There are few important factors to consider:

- The briefing phase should be kept as short as possible
- Providing the user with suggestions on potential interaction methods with the system should be avoided. Where techniques, such as paper prototyping [31] or an obvious *"Wizard of OZ"* [32] are used, providing a short explanation to avoid confusion is recommended.
- Important: It is imperative to inform the user that there are no incorrect interaction methods. User interaction is not being tested, but the system in which they are interacting with. Any occurring problems thus stem from flaws in the system design, not from their lack of proficiency.
- *"Think aloud"* [30]: The participants should be sensitized to articulate every single thought aloud during their interaction. This allows easy tracing of their thought processes over the interaction sequences.

3.3 II. Execution

Exploration Providing a short exploration phase prior to evaluating the actual task is recommended. This phase should remain short and provide the user with an opportunity to adjust with confronting a new system, as they would in a real scenario.

Testing While there are different ways to test multimedia information systems, a user test need not be purposeful to achieve prescribed goals. It can also be performed in an exploratory way by the user without a specific aim. Nevertheless, in scientific literature, an effective and gap-less user test is based on different use cases [33] and scenarios. These scenarios have to be created initially, which the study participant subsequently experiences. The user will be requested to follow certain tasks, and or, achieve objectives. The user will also be requested to use the *"think aloud method"*, where the study leader will record the attributes mentioned based on the user's current experience.

3.4 III. Questioning

Free Talk / Reflection Once the participant has completed all the required tasks, encouraging the user to discuss their overall experience is highly recommended instead of starting the questioning immediately. It is important to give the user an opportunity to discuss their mindset , before questions steer the user in a specific direction.

Questions Some users may have failed to take initiative in offering their insights. It may be relevant to elicit attributes by questioning the user about particular moments in which joy and discomfort was experienced. This aspect is crucial to facilitate discussion about their experience, namely their perceptions and feelings. Furthermore, questions about specific user decisions should be asked to prevent confusion. Where necessary, a small set of questions concerning which features require feedback, can be prepared beforehand.

3.5 IV. Assembling Item List

Collecting Attributes During the questioning process, the participants will have used different attributes to describe their experiences. For example: "too speedy, colorful, very funny, not understandable, creative, exciting...". These attributes should be listed. Moreover, the participants can be asked to describe their experience of the evaluated system in a post-questioning process. These adjectives would also need to be added to the item list.

Inverting Attribute List Following the interviews with all the study participants, the attribute list should be fairly substantial by now. It is worth mentioning that we experienced the attributes list to be shorter when performing group interviews. New, creating an inverted item list by using the opposites of all attributes (e.g. "too speedy→ slow, colourful→drab, very funny→serious, not understandable→easy to understand, creative→uninspired, exciting→boring"), is the next and most crucial step.

3.6 V. Results Gathering

Anonymous Rating The next stage requires the users to rate the inverted item list. It is important to communicate that the performed rating will be treated anonymously. (Requesting the user's gender, age, and other relevant data is still acceptable and should be done.) For the rating, a 7-point Likert scale [34] appears suitable, but other scales may fit also [35]. The study participants are now prepared to rate the system on attributes, which are actually the opposite of their own valuation. This may foster an increased critical reflection, since the attributes are contrary to their opinion (and mostly negative when they had a positive experience). This method provides the participants with an opportunity to reflect on a level with greater profundity.

Note: The process of compiling the inverted item list is time consuming, since it can only be completed after gathering attributes from several participants. If the participant is no longer physically present, the questionnaire can still be sent via a small online survey to the participant. However, long temporal gaps between completing the study and filling out the questionnaire should be avoided.

Analysis After all the results are collected, the overall average of each opposite attribute can be analyzed or converted into the original attribute. For example, the majority may have rated the attribute *"boring"* with a 3, The item be inverted into the original attribute, *"exciting"*, which would result in a rating of 5 (based on a 7-point Likert scale). The gathered data can now be analyzed by statistical means. Although ratings on a Likert scale is strictly spoken non-parametric data [36], a 7-point scale to account for parametric data is also suitable. Literature proposes that both are feasible [37], as the drawn conclusions may not necessarily differ [38].

4 Evaluating UGIIE

Based on past experiences, we found that study participants tended to feel increasingly annoyed when required to answer a large number of unsuitable questions from standardized methods. Literature has also shown goal inversion [4] to enable a more creative and impartial judgement and problem solving [5]. Therefore, UGIIE will potentially provide a more critical impression of the user experience and decrease user frustration during the post-questioning process

(*see Hypotheses*). To evidence this, the UGIIE will be benchmarked against established UX methods (*see Methodology*). Three experiments were designed, in which the UX of different products were measured. Finally, the results are presented and more general conclusions will be drawn (*see Results*).

4.1 Hypotheses

Hypothesis 1: UGIIE will yield a greater critical and consistent evaluation compared to other UX methods.

Hypothesis 2: UGIIE will apply a more relevant criteria for evaluation than standard UX questionnaire.

Hypothesis 3: UGIIE will be rated as less annoying compared to other standard questionnaires.

4.2 Methodology

To evidence the hypotheses, a comparative approach is used, in which different evaluation techniques are evaluated. Two established methods were used to compare their performances against the newly developed UGIIE method.

1. *User Experience Questionnaire (UEQ)*
2. *Modular Evaluation of Key Components of User Experience Questionnaire (meCUE)*
3. *User Generated Inverted Item Evaluation (UGIIE)*

4.3 Procedure

At the beginning of each experimental study, the participants were briefly introduced to the product. After completing the task, each user was asked to fill out a questionnaire in which they had to provide demographic data and rate the system with the help of a questionnaire. A drone, an unknown smartphone app, a projector, and the Instagram smartphone app were deliberately chosen from a broad range of interactive products. Four unique situations were examined and assessed: a group of users with no involvement in utilizing the product (drone), a group of users with great involvement in utilizing the product (unknown smartphone app), a group of users with a blend of limited understanding in utilizing the product (projector), and a group of users using a familiar smartphone app (Instagram). Experiment 4 was conducted remotely via a video-communication tool. At the end of each experiment, the users were requested to fill the survey questionnaires to measure the user experience of using the device. The users were also requested to mark the questions in the questionnaires which they found inappropriate. All four experiments were audio-taped with the participant's consent.

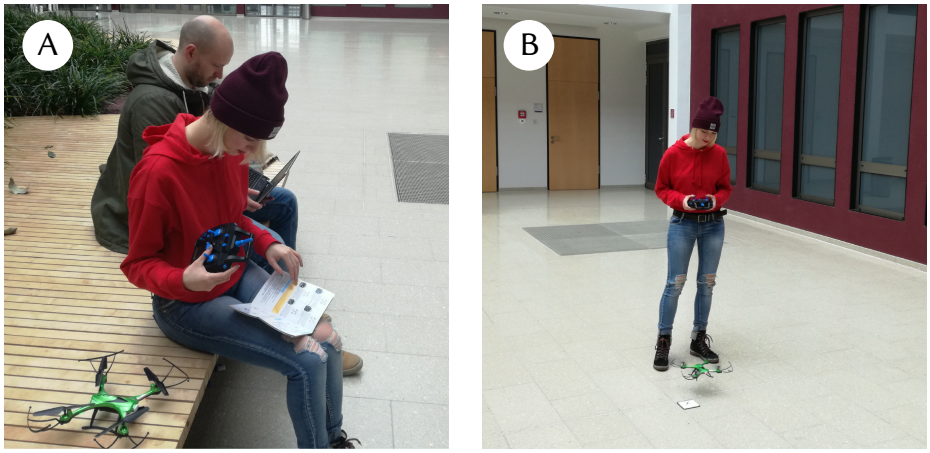

Fig. 2. Study participant getting familiar with the drone using the manual (A) and eventually performing the given tasks (B).

4.4 Task 1: Drone

The participants were seated at a table, while having a drone and wireless controller laid in front of them. The study leader sat next to them and read the tasks aloud, which were recorded down on a sheet of paper. A secretary accompanied the study leader and noted the protocol of the user's mentioned attributes. Subsequently, the users were requested to envision themselves as a drone pilot and test the drone by performing a series of tasks. Users can read the user manual at the beginning of the experiment or between the experimental tasks. The first task is to raise the drone roughly at the user's eye level and land it instantly on the marked position. The second task requires lifting the drone roughly to

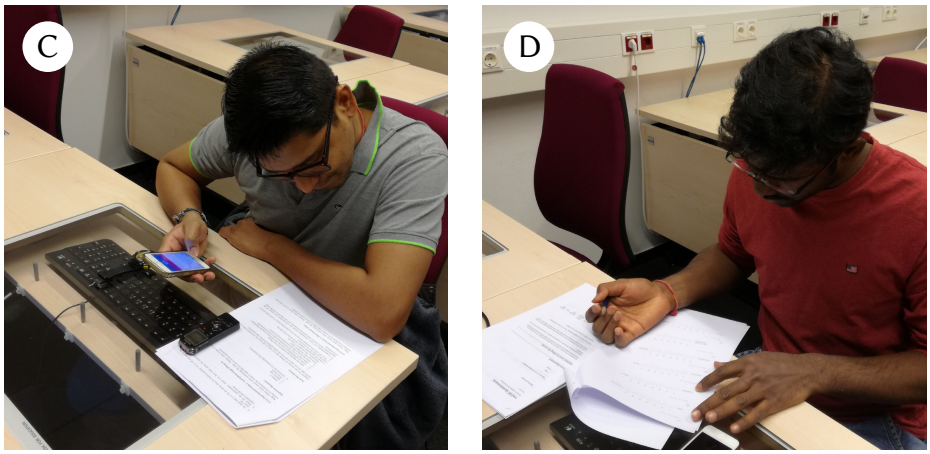

Fig. 3. Study participants performing the task on the smartphone (C) and filling out a questionnaire after the experiment (D)

the user's knee level and move the drone to hit the target (empty water bottle). The user will then land the drone immediately next to the target. The final task requires the user to fly the drone from one marked position to another.

4.5 Task 2: Unknown Smartphone App

The participants were seated at a table, while having a smartphone (model: iPhone 5) laid in front of them. The study leader sat next to them and read the tasks aloud, which were recorded on a sheet of paper. Beside the study leader, a secretary was taking protocol of the user's mentioned attributes. The users are then requested to envision themselves in a scenario where they need to create a sale advertisement to sell an old wristwatch using a specific mobile app they were not familiar with. Several sub-tasks, such as taking a photo of watch, were included.

4.6 Task 3: Projector

The participants were seated at a table, while having a mini projector (UNIC UC 46+) laid in front of them. The study leader sat next to them and read the tasks aloud, which were recorded on a sheet of paper. Beside a study leader, a secretary was taking protocol of the user's mentioned attributes. The users are requested to envision a scenario in which, during their thesis defence, the projector in the conference room experienced some issues. Subsequently, they were provided with a small portable projector to complete the thesis presentation. Users can read the user manual at the beginning of the experiment or between the experimental task. The first task is to connect the mini projector to the power supply and modify the setting to get a clear image output. The second task is to connect the laptop with the mini projector to project the thesis presentation slides onto the

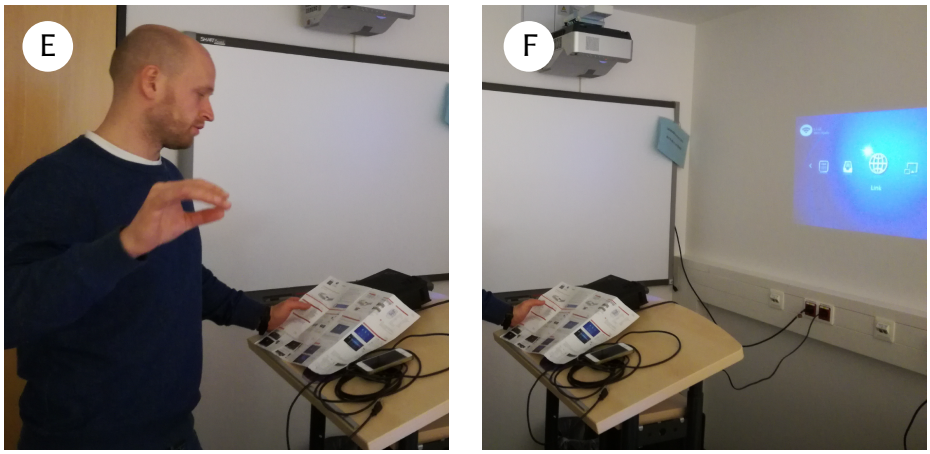

Fig. 4. Study participant performing the think-aloud technique during the experiment (E) and setting up the projector (F).

screen. The third task is to connect the projector to the smartphone by remotely using WiFi.

4.7 Task 4: Familiar Smartphone App

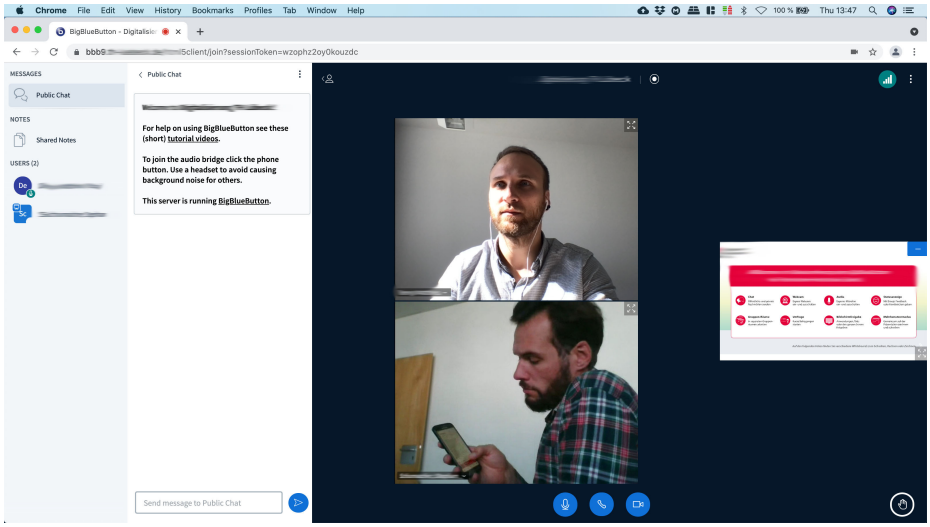

Fig. 5. Experimenter observing the study participant interacting with their smartphone during experiment 4.

Since this experiment was carried out during the COVID-19 pandemic, the study participant was located in a remote location from the experimenter. A pre-condition for this study was that participants were an Instagram user. The participant used their own Android smartphone for the study. For the first task, the user was required to set up the camera from a side-angle to allow the experimenter to gain a greater impression of the participant's performance. The experimenter guided the user through several use cases, which included: taking photos and uploading them; taking short videos and uploading them; searching the experimenters Instagram account and adding him as a friend; interacting with the experimenter via Instagram's chat, like, and comment feature. Finally, the experimenter asked the participant to customize their account, such as filling in an account description.

4.8 Participants

To perform the comparative evaluation of the three UX assessment techniques, 21 participants were invited to take part in the first three experiments. One participant was excluded afterwards. All participants were university students. The majority of the participants were male (85%) and the remaining were female (15%). The age of the participants ranged between 20yrs to 47yrs (M=25.7yrs).

The fourth study also featured 20 participants aged between 20yrs to 47yrs (M=24.55yrs). Since we were interested in comparing UX techniques and not the products itself, we neglected an age and gender balancing.

4.9 Data Gathering

As the UGIIE questionnaire needs to be generated during the experiment, we asked the first 10 participants to fill out the UEQ and meCUE questionnaires in alternating order. We decided not to ask participants to fill out more than two questionnaires to prevent unnecessary annoyance, as this could negatively bias the provided data from a third questionnaire.

To ensure comparability of all three questionnaires (UEQ, meCUE, and UGIIE), the rating scale was transformed to a common scale. All participants were asked to rate each attribute for all UX questionnaires using a 7-point Likert scale [34]. The following rating was applied: Strongly Disagree (1) to Strongly Agree (7). With each question, the participant could indicate whether the question was inappropriate and inapplicable. Although when the question was marked as such, the participant was still required to answer it regardless.

One participant that performed experiment 1-3 was declared an outlier since he provided unrealistic and incomplete answers. After removing this single dataset, we accumulated 30 valid questionnaires for each experiment, x10 meCUE, x10 UEQ, and x10 UGIIE, resulting in 120 filled questionnaires in total.

To gather the items for the UGIIE method, the mentioned attributes were extracted from the audio tapes. Noting these attributes during the experiment would be less time consuming. The analysis of the audio tapes resulted in the identification of 17 items for experiment 1 *(see Table 1 & 2)*, 14 item for experiment 2, 15 attributes for experiment 3, and 15 attributes for experiment 4.

4.10 Results

Addressing Hypothesis 1 Following the hypothesis, the rating of a system using a standardized posteriori questionnaire, may not be sufficiently critical or very reflective of the actual user experience. To investigate this, the ratings of common attributes, which were used for all three UX methods, were compared *(Common items – see Table 1 & 2)*.

Experiment 1 - Drone, six common attributes were identified: Easy, Understandable, Annoying, Attractive, Clear, Exciting.

Experiment 2 - Smartphone (unfamiliar App), four common attributes were identified: Easy, Attractive, Clear, Good.

Experiment 3 - Projector, four common attributes were identified: Understandable, Easy, Good, and Attractive.

Experiment 4 - Projector (Instagram), four common attributes were identified: Attractive, Friendly, Easy, and Practical.

Data Analysis: Although statistical significance is lacking among most single attributes (*see Figure 6*), a *one-way ANOVA for independent samples* could evidence a statistical significance for a couple of attributes.

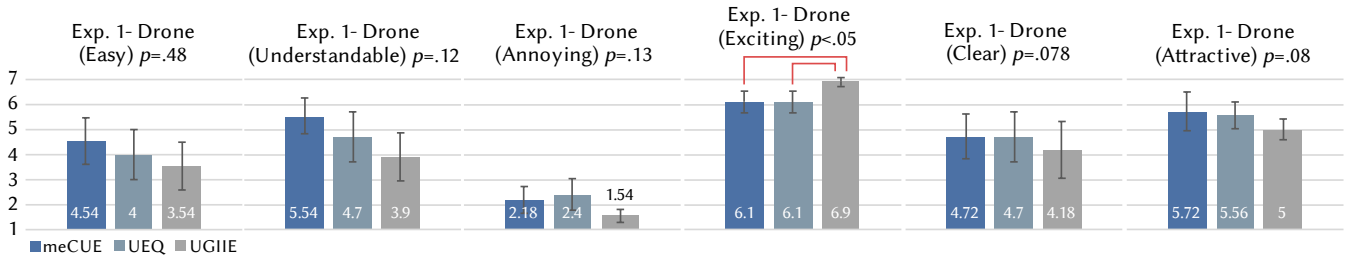

Fig. 6. Displaying all common attributes among the three questionnaires (UEQ, meCUE, UGIIE). UGIIE demonstrated to be somewhat closer to the baseline: the drone was rated to be very exciting.

In experiment 1, every participant found it very "exciting" flying the drone, which UGIIE (M=6.9; SD=.31) emphasized significantly stronger ($F_{2,27}$=3.92; p<.05) than meCUE (M=6.1; SD=.87) and UEQ (M=6.1; SD=.87), as confirmed by a Tukey HSD post-hoc analysis. For the other common attributes, it could be said that UGIIE emphasizes the user experience slightly stronger than other questionnaires. When there was a positive trend, UGIIE showed more positive results. When there is a negative trend, UGIIE showed a slightly more negative trend. However, these observations at all other attributes are not backed with a statistical difference (p>.05).

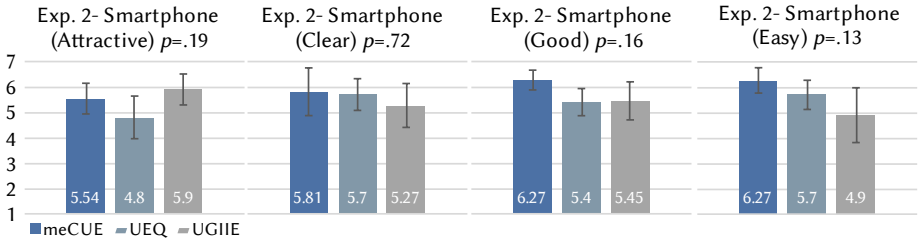

Fig. 7. Displaying all common attributes among the three questionnaires (UEQ, meCUE, UGIIE). UGIIE did not show a significant different rating with the first smartphone experiment.

In experiment 2, all common attributes showed consistency across all three questionnaires (p>.05), see Figure 7.

In experiment 3, the user had to setup a poorly designed low-budget WiFi projector. None of the participants were satisfied, as they stated during the experiment that the product had a bad output, was poor quality, and operated insufficiently. The participants were hardly able to accomplish the given task. Taking this as a baseline, the scores of the meCUE (M=4.6; SD=1.78) and UEQ (M=4.9; SD=1.45) do not reflect these "bad" experiences. Those tests score a neutral to positive tendency. In contrast, UGIIE (M=2.5; SD=1.71) shows a significantly more critical rating, which is confirmed by a one-way ANOVA for independent samples $F_{2,27}$=6.26; p<.05, as well as a post-hoc analysis by

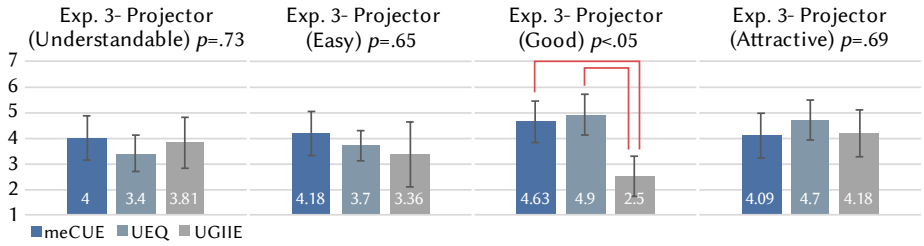

Fig. 8. Displaying all common attributes among the three questionnaires (UEQ, meCUE, UGIIE). UGIIE demonstrated to be somewhat closer to the baseline: the projector was perceived to be very bad).

a Tukey HSD. For other attributes UGIIE coincides with the other methods, although it is striking that UGIIE yields either similar ratings or more critical ratings closer to the qualitative statements. Considering the qualitative feedback during the studies as a ground truth baseline, the rating of all common attributes for both meCUE (M=.61; SD=.65) and UEQ (M=.68; SD=.58), deviated higher than the rating of UGIIE (M=.28; SD=.43). A statistical difference using a one-way ANOVA for independent samples $F_{2,39}$=2.02; p=.14 could not be indicated from this rather low sample size.

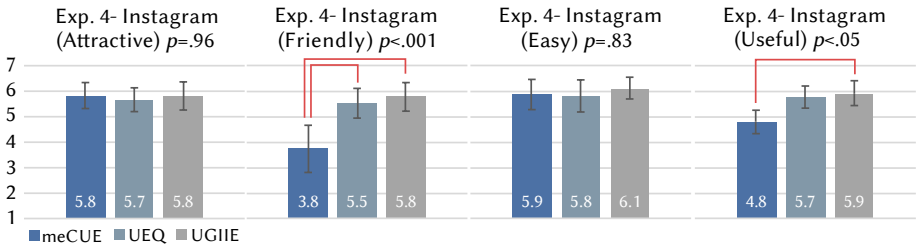

Fig. 9. Displaying all common attributes among the three questionnaires (UEQ, meCUE, UGIIE). UGIIE demonstrated to be somewhat closer to the baseline: the Instragram Smartphone app was perceived significantly different to meCUE results).

In experiment 4, we evaluated a sophisticated smartphone app (Instagram) optimized for great user experience. Every participant was positive about the app, as it was a requirement that participants were frequent users. For the attribute "friendly", meCUE (M=3.8; SD=1.8) provided a significant understatement compared to UEQ (M=5.5; SD=1.18) and UGIIE (M=5.8; SD=1.14), following a one-way ANOVA ($F_{2,27}$=5.85; p<.001) and a Tukey HSD post-hoc analysis (*see Figure 9*). Another difference was found in the item "useful". meCUE (M=4.8; SD=.92) again under-performed while UGIIE (M=5.9; SD=.99) showed a significantly better and more realisitc result, as it coincided with results gathered from the UEQ (M=5.7; SD=.82). A one-way ANOVA

($F_{2,27}$=4.1; p=.023) made the discovery, while a Tukey HSD test confirmed the result. Other common attributes, such as "attractive" and "easy" did not significantly differ (p>.05) and seemed constant across all three tests.

Considering qualitative as ground truth, (1) UGIIE was indicated to be slightly more critical at some attributes and even significantly different to four attributes, which also coincides with the user statements and, (2) the tendency of UGIIE to deviate less from the baseline compared to meCUE and UEQ, allows the conclusion that hypothesis 1 is acceptable.

Addressing Hypothesis 2 A simultaneous advantage and disadvantage of commonly applied UX methods, is the variety of the questions asked. These UX methods are advantageous as they can assess a variety of different systems. However, this also requires answering a large amount inappropriate questions, which may be perceived as bothersome and time consuming. In fact, the meCUE Questionnaire incorporates 33 questions, in which 22 were found to be inappropriate across all experiments. Therefore, only M=33.33% (SD=8.9) of all attributes were considered appropriate. The UEQ Questionnaire includes 26 attributes, where approximately 20 were considered unsuitable across all experiments. Thus, only M=24.35% (SD=4.79%) of all attributes were suitable. In contrast, the UGIIE accumulated, on average, 15 different attributes across all experiments. Less than two items were considered inappropriate across all participants. Therefore, M=88.41% (SD=8.83%) of all attributes were perceived as appropriate.

A one-way ANOVA for independent samples ($F_{3,6}$=42.5; p<.001) suggests a strong significance comparing all UX questionnaires. A Tukey HSD Test confirms that the UGIIE yields significantly more appropriate attributes compared to the meCUE (p<.01) and the UEQ (p<.01). In terms of appropriate attributes, there are no significant differences between the meCUE and the UEQ. Therefore, this data confirms hypothesis 2 in that UGIIE applies more relevant attributes than classical UX methods, which is due to the sheer nature of how a UGIIE questionnaire is developed.

Addressing Hypothesis 3 The participants were requested to think aloud during the experiment, as well as during the time when filling out the questionnaire. Commenting on the participant's reaction was avoided to prevent imbuing any biases. It was striking that most participants stated that several items in the standard questionnaires were unrelated. Comments from the participants are as follows:

P12: *"It [–the attribute to be rated–]] is not specific."* P10: *"[Several attributes were] out of Scope"* P2: *"It is too open, the questionnaire."* P17: *"I am unable to answer the questions, it does not reflect my real experience"* P21: *"It [–the question–] is too unspecific."* P1: *"It's difficult to answer the question which is not applicable."* P4: *"It is not related."* P14: *"[This question is] exaggerated."*

Another feedback from the questionnaire was the apparent excess of similar items. The majority of participants stated that they were irritated by the high similarity in questions, having felt they had addressed a similar question prior.

Once the participants took part in more than one experiment and used the UGIIE, the majority concluded that UGIIE was quick to fill out and did not include irrelevant questions. Based on the many qualitative feedback, hypothesis 3 can thus be confirmed, as the UGIIE method creates less annoyance compared to other standard questionnaires (UEQ and meCUE) that often show unsuitable questions and questions of high redundancy.

5 DISCUSSION

The proposed UX method: User Generated Inverted Item Evaluation (UGIIE) features certain advantages in contrast to standardized UX questionnaires. Still, drawbacks and challenges are an inevitable aspect of its development.

5.1 Advantages

Critical Reflection and Consistent Feedback Due to the poor usability of some products we tested, it was evident that the participants experienced difficulties. Their negative experiences were also confirmed by the attributes they used to describe their situation at that particular moment. However, these experiences were reflected differently with a standardized questionnaire. UGIIE's inverted item questionnaire, which is contrary to their own valuation, potentially enabled a more critical reflection. The results appeared more consistent, as they greater coincided with the users' actual qualitative feedback spoken aloud during the experiment. This finding is also somewhat underpinned by our subjective observation that most participants were slightly surprised when asked to fill in UGIIE. Participants also seemed more cautious when rating the items on UGIIE.

Appropriate Attributes Due to the nature of UGIIE, it includes more appropriate questions than standardized questionnaires. From the ascertained outcome, only 24.35% of these attributes in UEQ method were rated as appropriate, 33.3% of all attributes were considered proper in meCUE questionnaire, and 88.4% of all attributes were considered appropriate in UGIIE. Although standard questionnaires may not require a preparation time, the costs of creating UGIIE questionnaires are still relatively low when compiling it based on a small number of users, such as five. The benefits indicated from the results, namely more appropriate questions, reveal that UGIIE is a worthwhile investment.

Subjectively Less Annoying Questions In comparison to commonly used questionnaires, which usually deploy up to 33 items that may be partly redundant, UGIIE is significantly more compact. In over four experiments, we extracted 15 relevant items on average. The participant's qualitative discussion statements concerning the rated attributes of the questionnaires reveal that the UGIIE method was experienced as less annoying and less frustrating compared to the standard UX methods.

Broad Applicability The nature of UGIIE does not require it to be bound to a specific domain - *see Experiment 1-4*. It is applicable in a variety of scenarios, which do not necessarily need to be within interactive products. However, one still needs to rigorously experiment and prove its usefulness for a variety of other products and systems.

5.2 Further Insights

Extracting a Reasonable Number of Attributes Extracting items is an essential step for the UGIIE technique. It was identified that after five times running the experiment, at least 80% of the total attributes were discovered (based on our investigations with 10 runs). This suggests that very few participants sufficiently develop a reasonable and precise conclusion of the outcome.

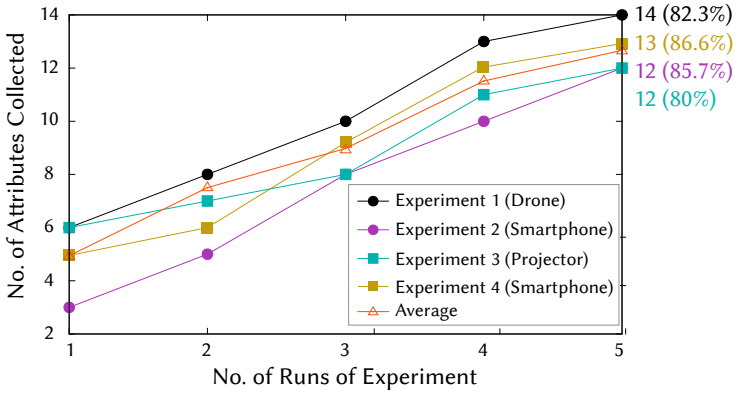

Fig. 10. Five runs seem to be sufficient in to extract a reasonable number of attributes. In our experiments, we could find at least 80% of the total attributes found after 10 runs.

Reasonable Time Consumption Compiling the tailored inverted questionnaire requires the experimenter to expend extra time. For instance, in our last experiment, we took around 2 hours to develop the questionnaire by fast-forwarding through the audio recordings of 10 users, although the actual use case test only lasted around 15 minutes per user. However, once compiled, the presence of the experimenter is no longer necessary. Also, the time participants require to fill the questionnaire is shorter compared to standardized questionnaires. This is advantageous when testing a greater number of users.

UEQ Comparison On the one hand, the User Experience Questionnaire (UEQ) provides a comprehensive impression of UX, ranging from classical usability aspects to user experience aspects. However, the six scale with the 26 items constrains the UX measurement to certain limited items. On the other hand, UEQ does not require significant reading efforts from the participants and can be completed within five minutes. Another advantage is that UEQ is of no cost.

meCUE Comparison MeCUE is based on statements and considers the affection of product use. With the meCUE questionnaire, the authors express that it has the advantage of assessing major UX components in a comprehensive manner. However, the questions are often redundant and lengthy, creating annoyances with the survey method. Nevertheless, meCUE may be relevant to researchers, as it is applicable to any User Experience surveys in a variety of interactive systems.

5.3 Challenges

Preparation Time Required The UGIIE method features a custom questionnaire, which requires preparation in advance. This creates a delay between the actual execution of the study and administering the posterori evaluation. To counter this, items from other users who are testing the system were collected first. The generated inverted questionnaire was then deployed straightaway with other users, once they completed using the system. Alternatively, the UGIIE questionnaire could be sent to the users some days after testing the system. Recapturing their UX with a time-distance, may enable a greater critical and consistent reflection. However, this also yields drawbacks, as the users may have forgotten their actual experience of the system during the execution. If just testing a single-digit number of users, UEQ and meCUE have a time-cost advantage.

Item Extraction can Cause Complications In some exceptional cases, problems may arise, namely when the system is tested with a passive user who is reserved in character. Although previously prompted to use the think-aloud technique, such a user may still fail to use their initiative to discuss their thought processes aloud. This creates difficulties in capturing the user's actual experience. Continuing to prompt the user to discuss their experience, may create user dissatisfaction and imbue a sense of bias. In this case, the chances of failing to capture a realistic UX may increase. Only a greater sample size of users can mitigate this issue.

5.4 Methodological Limitations

Evaluating Evaluation Tools The optimum method to evaluate an evaluation tool remains contentious. While a comparative approach may be superior, it also has limitations. An issue which arises is that each method uses their own rating scale. Moreover, a comparison is not necessarily fair, as different UX methods are tailored to specific domains. To ensure the best comparison possible, two similar UX questionnaires were selected for interactive systems. The same rating scale, a 7-point Likert scale, was applied.

Participants In reality, there is no ideal user. Instead, the individual factor is very well pronounced, which underpins much of UX research. Therefore, a high number of participants and trials are suggested to counterbalance the individual

factor, and thus enable generalized and valid conclusions. As each experiment (including: preparation, execution, filling the questionnaire) lasted around an hour, roughly around three hours per participant ($\sim$ 80 hours total study time), a reasonable cut-off point was necessary. This point was defined when a minimum of 10 questionnaires were completed for each method. Running statistical means with an increased number of participants, may have contributed to more significant results and possibly fully proving the first hypothesis.

Ground Truth UX As there is no physiological technology to measure the actual UX, we considered qualitative feedback expressed during the study as the ground truth UX, which we compared with the quantitative ratings.

6 Related Work

6.1 Overview of UX evaluation methods

There are various User Experience evaluation methods within the market. A few common ones are: AttrakDiff, Experience Sampling Method (ESM), Game experience questionnaire (GEQ), User Experience Questionnaire (UEQ),TRUE Tracking Realtime User Experience , Modular Evaluation of Key Components of UX (meCUE) questionnaire, Attrak-Work questionnaire etc. Several overviews of UX assessment techniques have been conducted [39]. For instance, Isomursu et al. [40] particularly highlights those surveying transient feelings. [41] lists a collection of more extensive methods for designing pleasurable products. Vermeeren et al. arranges UX evaluation strategies into three gatherings, namely by the type of measures the method focuses on: Sensory characteristics, Articulation or Meaning and Emotional response [3]. The AAAC [42] gathered and developed design and assessment techniques for effective interactive systems. Furthermore, Isomursu [43] arranged alternative UX assessment techniques that focus on understanding user emotions.

However, the broadest collection includes 96 UX assessment techniques [44]. Based on the current individual UX evaluation strategies [3], inferences can be drawn as to which techniques are rare or successful, and what their qualities and shortcomings are. This overview is briefly summarized [3] by using eight criteria, based on their value in interest:

Origin: It is striking that 70% of all strategies root in the scholar community. Still, it is plausible that a large number of UX assessment strategies, categorized by industry, remain undiscovered.

Type of Gathered Information: Around 33% of the strategies were accounted for giving quantitative information, 33% qualitative information and 33% both.

Type of Applications/Designs: 73% of the strategies are moderately application-autonomous. However, most UX techniques aim to evaluate: web services, mobile software, PC software, and hardware designs.

Data Sources: The majority (80%) can be utilized with single users. 17% use a user groups, as a conceivable source of data (e.g., AttrakWork questionnaire, outdoor play observation scheme, Living Lab, product personality assignment).

Location: Lab, field, or on the web? About half of all strategies are only applicable at a single location: in the lab (67%), in the field (52%), or on the web (40%). All other UX evaluation strategies can be applied in multiple contexts and locations.

Period of Experience: 63% of all strategies can be utilized for considering UX of single behavioural episodes and 59% are suggetsed to be applied in a single test session. 36% of all techniques can manage long-term use.

Development Phases: 39% of all methods are to be used to assess the system's UX in an early development stage. However, the vast majority of all strategies (∼80%) can be utilized as a part for later as well as early development stages.

Technical Requirements: Most techniques (67%) are accounted for not requiring any unique equipment or programming. Remote utilization, such as by means of a website is conceivable in about half of all cases (e.g., multiple sorting methods, ServUX, audio narrative, activity experience sampling, SUMI, etc.)

From the above UX evaluation framework, User Experience Questionnaire (UEQ) and a Modular Evaluation of Key Components of UX (meCUE Questionnaire) provide exceptional advantages, as similarities exist with our proposed method. UEQ and meCUE both provide a complete UX impression, ranging from classic usability aspects to user experience aspects. Moreover, it introduces an analytic apparatus to precisely translate the outcome effectively. UEQ and meCUE are also highly accessible, being of no cost [45]. Therefore, comparing UGIIE with these established UX questionnaires may produce interesting results.

6.2 Quantitative UX Evaluation Methods for Single Episodes

Quantitative evaluations represents the user's subjective feelings towards the item they utilize in a measurable way enabling a comparison. Estimating a valid impression on UX the user encounters with a design or service generally requires gathering input from a large user group. The most efficient way is using an online survey [45]. Common UX techniques are mostly questionnaire-based, which provide a quantitative analysis. As previously stated, the User Experience Questionnaire (UEQ), and a Modular Evaluation of Key Components of UX (meCUE), were selected based on their advantages and high similarities to our proposed method.

User Experience Questionnaire (UEQ) The principal objective of the UEQ is to enable a prompt estimation of the user experience of interactive products [46]. The user experience questionnaire contains six scales with 26 items in total (Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, Novelty) [45]. UEQ features a bi-polar item scale, which can be rated on seven steps. For example, one item would be annoying <> enjoyable. This rating scale has been benchmarked [47] with data sets containing information from a 163 item assessment. These assessed items secured an extensive variety of applications, such as: complex business applications (98), advancement tools (4), web shops or services (37), social networks (3), versatile applications (13), and a few other

products(8). UEQ has also been applied in research context and counts to a common technique to measure UX [48].

meCUE - A Modular Evaluation of Key components of UX The meCUE questionnaire [49] is described by its creators as *"a freely accessible, experimentally established questionnaire, which centers around the particular securing of user focused audits and their experience of interactive technical product"* [50]. It is a particular UX evaluation scale adapted from the Thüring and Mahlke's CUE- model [51] and composed of 33 items partitioned into 4 dimensions: instrumental and non-instrumental product perception, emotions, consequence and overall judgment. Inside each sub-scale, respondents are requested to evaluate their agreement level with proclamations on a 7-point Likert scale [34] from 1 "strongly disagree" to 7 "strongly agree". All questions are positively worded and required, as expressed in the instructions. As indicated by its' authors, the main advantage of meCUE when contrasted with existing questionnaire is to survey the significant segments of UX in a comprehensive manner. The psychometric properties of the questionnaire have been evaluated through a few investigations [49]. Furthermore, *"meCUE can be applied in UX reviews on a wide range of interactive systems"* [50]. The meCUE questionnaire has recorded logical properties, particularly taking into account the evaluation of both hedonic and pragmatic perspectives and is depicted as reasonable for all specific circumstances. Using a more *"common-sensual"* decision from a pragmatic-only scale is not ideal for the present purpose [50].

7 Conclusion

This paper presented an alternative UX evaluation technique, the User Generated Inverted Item Evaluation (UGIIE). The core idea is adapted from reverse brainstorming, in which the user's mind is freed from thought processes bound by limiting grids of goal inversion. Based on the think-aloud technique and a post-questioning method, only the relevant attributes that describe the system were evaluated and later inverted, before being subject to participant ratings. By comparing UGIIE against two common UX questionnaires (UEQ and meCUE), we discovered that in some circumstances, UGIIE was rated as being significantly more basic for a few traits. This coincides with the subjective user explanation and the general factual contrast of UGIIE compared to a baseline average calculated from the UEQ and meCUE. Moreover, a UGIIE questionnaire incorporates more relevant questions and is thus considered to be less annoying for the participant to complete. Incidentally, this positively impacts the UX of the evaluation process itself. Although standard questionnaires may not require preparation time, applying UGIIE becomes increasingly advantageous when testing greater number of users. However, aiming to gather quick insights on UX with just a single digit number of users, standard questionnaires or qualitative questioning seem to provide the most time-efficient solution.

References

1. C. Lallemand, G. Gronier, and V. Koenig, "User experience: A concept without consensus? exploring practitioners' perspectives through an international survey," *Computers in Human Behavior*, vol. 43, pp. 35–48, 2015.
2. I. FDIs, "9241-210: 2009. ergonomics of human system interaction-part 210: Human-centered design for interactive systems (formerly known as 13407)," *International Organization for Standardization (ISO). Switzerland*, 2009.
3. A. P. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: current state and development needs," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 521–530, ACM, 2010.
4. C. Wilson, *Brainstorming and beyond: a user-centered design method.* Newnes, 2013.
5. T. Rickards, *Problem-solving through creative analysis.* Wiley, 1974.
6. N. Corporation, "Inspired human technology," 2005.
7. M. Seidel, C. H. Loch, and S. Chahil, "Quo vadis, automotive industry? a vision of possible industry transformations," *European Management Journal*, vol. 23, no. 4, pp. 439–449, 2005.
8. K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl, "Towards practical user experience evaluation methods," *Meaningful measures: Valid useful user experience measurement (VUUM)*, pp. 19–22, 2008.
9. M. Hassenzahl, "User experience (ux): towards an experiential perspective on product quality," in *Proceedings of the 20th Conference on l'Interaction Homme-Machine*, pp. 11–15, ACM, 2008.
10. E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 719–728, ACM, 2009.
11. S. Kujala and K. Väänänen-Vainio-Mattila, "Value of information systems and products: Understanding the users' perspective and values," *Journal of Information Technology Theory and Application (JITTA)*, vol. 9, no. 4, p. 4, 2009.
12. E. Karapanos, J. F. Zimmerman, and J. Forlizzi, "J., & martens, j.(2009). user experience over time: an initial framework," in *Proceedings of the 27th international conference on Human factors in computing systems*, pp. 729–738, 2009.
13. S. Kujala, V. Roto, K. Väänänen-Vainio-Mattila, E. Karapanos, and A. Sinnelä, "Ux curve: A method for evaluating long-term user experience," *Interacting with Computers*, vol. 23, no. 5, pp. 473–483, 2011.
14. B. Johanson, A. Fox, and T. Winograd, "The interactive workspaces project: Experiences with ubiquitous computing rooms," *IEEE pervasive computing*, vol. 1, no. 2, pp. 67–74, 2002.
15. L. Arhippainen and M. Tähti, "Empirical evaluation of user experience in two adaptive mobile application prototypes," in *MUM 2003. Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia*, no. 011, pp. 27–34, Linköping University Electronic Press, 2003.
16. J. Dewey, "Art as experience: New york: Gp putnam's sons," 1980.
17. J. Forlizzi and S. Ford, "The building blocks of experience: an early framework for interaction designers," in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pp. 419–423, ACM, 2000.

18. W. Albert and T. Tullis, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.
19. E. L.-C. Law, "The measurability and predictability of user experience," in *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*, pp. 1–10, ACM, 2011.
20. J. A. Bargas-Avila and K. Hornbæk, "Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2689–2698, ACM, 2011.
21. W. Iso, "9241-11. ergonomic requirements for office work with visual display terminals (vdts)," *The international organization for standardization*, vol. 45, p. 9, 1998.
22. V. E. ROTO, "User experience white paper," *http://www. allaboutux. org/uxwhitepaper*, 2011.
23. N. Bevan, J. Carter, and S. Harker, "Iso 9241-11 revised: What have we learnt about usability since 1998?," in *International Conference on Human-Computer Interaction*, pp. 143–151, Springer, 2015.
24. M. Hertzum, "Images of usability," *Intl. Journal of Human–Computer Interaction*, vol. 26, no. 6, pp. 567–600, 2010.
25. M. Hassenzahl, "The effect of perceived hedonic quality on product appealingness," *International Journal of Human-Computer Interaction*, vol. 13, no. 4, pp. 481–499, 2001.
26. J. Kaye, "Evaluating experience-focused hci," in *CHI'07 extended abstracts on Human factors in computing systems*, pp. 1661–1664, ACM, 2007.
27. M. Q. Patton, *How to use qualitative methods in evaluation*. No. 4, Sage, 1987.
28. J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pp. 206–213, ACM, 1993.
29. M. Schmettow, "Sample size in usability studies," *Commun. ACM*, vol. 55, no. 4, pp. 64–70, 2012.
30. M. Van Someren, Y. Barnard, and J. Sandberg, "The think aloud method: a practical approach to modelling cognitive," 1994.
31. C. Snyder, *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann, 2003.
32. N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies—why and how," *Knowledge-based systems*, vol. 6, no. 4, pp. 258–266, 1993.
33. A. Cockburn, "Writing effective use cases, the crystal collection for software professionals," *Addison-Wesley Professional Reading*, 2000.
34. R. Likert, "A technique for the measurement of attitudes.," *Archives of psychology*, 1932.
35. T. J. Maurer and H. R. Pierce, "A comparison of likert scale and traditional measures of self-efficacy.," *Journal of applied psychology*, vol. 83, no. 2, p. 324, 1998.
36. G. Norman, "Likert scales, levels of measurement and the "laws" of statistics," *Advances in health sciences education*, vol. 15, no. 5, pp. 625–632, 2010.
37. G. M. Sullivan and A. R. Artino Jr, "Analyzing and interpreting data from likert-type scales," *Journal of graduate medical education*, vol. 5, no. 4, pp. 541–542, 2013.
38. J. Murray, "Likert data: what to use, parametric or non-parametric?," *International Journal of Business and Social Science*, vol. 4, no. 11, 2013.
39. ENGAGE, "Report on the evaluation of generative tools and methods for 'emotional design'," *Deliverable D15.3 EU project Engage 520998*, 2006.

40. M. Isomursu, "User experience evaluation with experimental pilots," *Proc. of UXEM, www. cs. tut. fi/ihte/CHI08_workshop/papers. shtml*, 2008.
41. P. Jordan, "Designing pleasurable products," *An introduction to the new human*, 2000.
42. H. D9j, "Final report on wp9," 2008.
43. M. Isomursu, M. Tähti, S. Väinämö, and K. Kuutti, "Experimental evaluation of five methods for collecting emotions in field settings with mobile applications," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 404–418, 2007.
44. V. Roto, M. Lee, K. Pihkala, B. Castro, A. Vermeeren, E. Law, K. Väänänen-Vainio-Mattila, J. Hoonhout, and M. Obrist, "All about ux - information for user experience professionals," 2010.
45. H. B. Santoso, M. Schrepp, R. Isal, A. Y. Utomo, and B. Priyogi, "Measuring user experience of the student-centered e-learning environment.," *Journal of Educators Online*, vol. 13, no. 1, pp. 58–79, 2016.
46. B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*, pp. 63–76, Springer, 2008.
47. M. Schrepp, S. Olschner, and U. Schubert, "User experience questionnaire (ueq) benchmark. praxiserfahrungen zur auswertung und anwendung von ueq-erhebungen im business-umfeld," *Tagungsband UP13*, 2013.
48. J. Hartmann, "User experience monitoring: Über die notwendigkeit geschäftskritische online-prozesse permanent zu überwachen," *i-com Zeitschrift für interaktive und kooperative Medien*, vol. 10, no. 3, pp. 59–62, 2011.
49. M. Minge and L. Riedel, "mecue-ein modularer fragebogen zur erfassung des nutzungserlebens.," in *Mensch & Computer*, pp. 89–98, 2013.
50. C. Lallemand and V. Koenig, "How could an intranet be like a friend to me?: Why standardized ux scales don't always fit," in *Proceedings of the European Conference on Cognitive Ergonomics 2017*, pp. 9–16, ACM, 2017.
51. M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in human–technology interaction," *International Journal of Psychology*, vol. 42, no. 4, pp. 253–264, 2007.

APPENDIX

Table 1. Displaying the collected attributes and the corresponding inverted items for the first two experiments.

No.	User Generated	Inverted Item	No.	User Generated	Inverted Item
1	very sensitive	→ insensitive	1	too small	→ big
2	lightweight	→ heavy	2	good appearance	→ bad
3	not understandable	→ understandable	3	little bit slow	→ fast
4	unclear	→ clear	4	no direct feedback	→ direct feedback
5	hard to accomplish	→ easy to accomplish	5	dislike	→ like
6	uncontrollable	→ controllable	6	unclear	→ clear
7	problematic	→ straightforward	7	not responsive	→ responsive
8	very cheap	→ expensive	8	attractive	→ unattractive
9	attractive	→ unattractive	9	easy to find	→ difficult to find
10	confusing	→ simple	10	easy to use	→ difficult to use
11	complicated	→ uncomplicated	11	high quality	→ inferior
12	responsive	→ unresponsive	12	user friendly	→ not user friendly
13	colourful	→ colourless	13	understandable	→ not understandable
14	fun	→ boring	14	confusing	→ simple
15	enjoyable	→ annoying			
16	not easy	→ easy			
17	hard to learn	→ easy to learn			
1	**Drone**		**2**	**Smartphone**	**(unfamiliar App)**

Table 2. Displaying the collected attributes and the corresponding inverted items for the last two experiments.

No.	User Generated	Inverted Item	No.	User Generated	Inverted Item
1	noisy	→ quiet	1	attractive	→ unattractive
2	unclear (UI)	→ understandable	2	clear	→ unclear
3	not interesting	→ interesting	3	featured	→ featureless
4	confusing	→ clear	4	aesthetic	→ unsightly
5	bad output	→ good output	5	intuitive	→ abstract
6	fast	→ slow	6	intelligent	→ non-intelligent
7	insufficient	→ sufficient	7	fast	→ slow
8	difficult	→ easy	8	relaxing	→ stressful
9	poor quality	→ good quality	9	pleasant	→ unpleasant
10	small and portable	→ big and unhandy	10	interesting	→ dull
11	attractive	→ unattractive	11	friendly	→ unfriendly
12	useful	→ not useful	12	customized	→ non-modifiable
13	difficult to navigate	→ easy to navigate	13	easy to use	→ difficult to use
14	frustrating	→ pleasuring	14	useful	→ useless
15	challenging	→ not demanding	15	convenient	→ inconvenient
3	**Projector**		**4**	**Smartphone**	**(Instagram)**