

A Bermuda Triangle? - A Review of Method Application and Triangulation in User Experience Evaluation

Ingrid Pettersson*
Volvo Cars and Chalmers
University of Technology
Göteborg, Sweden
ingrid.pettersson@volvocars.com

Florian Lachner*
LMU Munich
Media Informatics Group
Munich, Germany
florian.lachner@ifi.lmu.de

Anna-Katharina Frison*
Technische Hochschule
Ingolstadt and Johannes
Kepler University Linz
anna-katharina.frison@thi.de

Andreas Riemer
Technische Hochschule
Ingolstadt and Johannes
Kepler University Linz
ansreas.riener@thi.de

Andreas Butz
LMU Munich
Media Informatics Group
Munich, Germany
butz@ifi.lmu.de

ABSTRACT

User experience (UX) evaluation is a growing field with diverse approaches. To understand the development since previous meta-review efforts, we conducted a state-of-the-art review of UX evaluation techniques with special attention to the triangulation between methods. We systematically selected and analyzed 100 papers from recent years and while we found an increase of relevant UX studies, we also saw a remaining overlap with pure usability evaluations. Positive trends include an increasing percentage of field rather than lab studies and a tendency to combine several methods in UX studies. Triangulation was applied in more than two thirds of the studies, and the most common method combination was questionnaires and interviews. Based on our analysis, we derive common patterns for triangulation in UX evaluation efforts. A critical discussion about existing approaches should help to obtain stronger results, especially when evaluating new technologies.

CCS Concepts

•**Human-centered computing** → **User studies; Empirical studies in HCI; HCI theory, concepts and models;**

Author Keywords

User experience; UX; evaluation; triangulation; mixed methods; review; meta-analysis

* First three authors contributed equally

INTRODUCTION

User experience (UX) has attracted increasing interest in recent years. One comparable indicator, at least for academia, is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

© 2018 ACM. ISBN 978-1-4503-5620-6/18/04...\$15.00

DOI: <https://doi.org/10.1145/3173574.3174035>

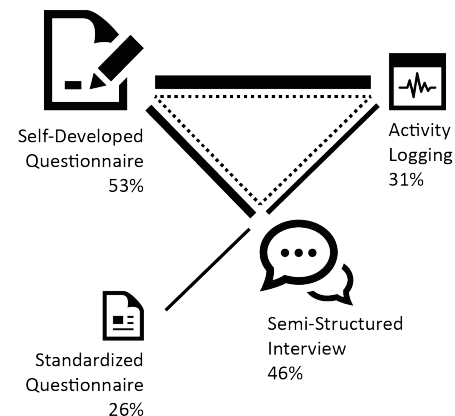


Figure 1. Commonly applied methods and triangulation strategies found in our literature study. The size of the nodes (methods) and links (combinations) is proportional to the number of occurrences.

the number of papers published. In Google Scholar, for example, the search term “user experience” roughly returns 20,800 results for the publication year 2010, increasing to more than 32,500 for the year 2016 (+56% in total). While UX considerations have become embedded in research and design processes, they still remain a challenging and strongly discussed area for both researchers in academia and practitioners in industry.

The focus of UX, for example defined as “A person’s perceptions and responses that result from the use or anticipated use of a product, system or service” [21], has shifted from simply acknowledging usability and performance aspects of interactive products, towards the emotional, hedonic aspects of interaction. Reaching beyond usability, UX has been extended to incorporate hedonic qualities such as aesthetics, stimulation and identification [37]. UX also is a dynamic concept influenced by contextual aspects, such as place, social and temporal aspects of use as well as the users’ specific emotional states [47, 23]. It is thus clear that UX offers a much more holistic and dynamic take on interaction with products than pure usability.

As an academic discipline, UX has evolved in the multi-disciplinary intersection of fields such as cognitive science, design, psychology and engineering. Not only the definition of the term UX is discussed, but also the question how to study and evaluate UX. A number of surveys of available methods and methodological gaps were conducted around the turn of the present decade (see, for example, [90, 107, 80, 9]). However, as the field has evolved over the years, we see the need to reassess the field of UX evaluation. We want to take the thoughts of the previous meta-analyses of Bargas-Avila and Hornbæk [9], Vermeeren et al. [107], Roto et al. [90] and Obrist et al. [80] - published around 2010 - further, to examine the progress until today. Our goal is to update the knowledge and provide an analysis of the current characteristics of empirical studies in UX from 2010 to 2016, as well as an outlook on possible future developments. In particular, we are interested in how the multi-dimensionality of UX is approached by using (or not using) method triangulation, i.e., applying two or more methods, to obtain well-founded evaluation results from different perspectives. Our analysis therefore includes trends regarding the number of UX publications, UX dimensions studied, study contexts, method application, and triangulation patterns derived from the analysis of method application.

Building on the current state of the art, we point out potential gaps and future directions. Is the topic of UX evaluation still a Bermuda Triangle of disparate research approaches, or is there land in sight for an evolved UX evaluation practice?

RELATED WORK

Evaluation has been identified as one of the core pillars of academic UX research [58, 63, 103]. Alves et al. [5] demonstrated in 2014 that experience evaluation plays an important role for UX practitioners in industry and concludes that “most practitioners believe that UX evaluations have a strong to decisive impact on the user interface”. The approaches used for empirical UX evaluation have, however, been debated (see for example [13, 63, 90]) as a result of the different epistemological directions of the research area.

It has been claimed that evaluating UX requires new assessment methods and approaches [9, 50] and a wide range of UX evaluation methods have been developed over the years (e.g., UX curve [56], AttrakDiff [39], or the UX-q [61]). Bargas-Avila and Hornbæk [9] pinpoint three areas of intense debate regarding evaluation: the types of products studied, the dimensions of UX, and the methodologies applied, reigniting “the debate between qualitative and quantitative approaches” [9]. Law et al. [63] point out the epistemological differences between engineering approaches to UX where “to measure is to know” and the humanities, where it may be considered naive and simplistic to assume that a fuzzy concept such as experience can be readily reduced and measured.

The ambition of sound empirical evaluations [63] applicable to UX thus led to the meta-review efforts in the late 2000’s and early 2010’s. These reviews concerned both the underlying theory and models of UX [58, 64] as well as the area of UX evaluation [5, 9, 80, 90, 107]. A number of common methodological gaps in UX evaluation approaches were identified at this point of time:

- In previous work, the underlying, assumed dimensions of UX have been found to often be unclear and/or interchangeable with usability [64]. The often very vague link between evaluations and UX theory is problematic; not fully understanding and formulating what is evaluated makes improvements difficult.
- Existing evaluation work was largely founded on self-defined questionnaires and few UX methods were satisfactorily validated for the cause [9, 107].
- There were only very few practices for evaluating UX at early stages of design processes [90, 107, 9]. These need formative methods, providing feedback that can provide information on what and why to revise to improve a design already early in the development [107].
- There was a lack of UX method triangulation that addresses the multidimensionality of experience. The lack of rational ways of collecting data from multiple methods was found by several reviews [9, 90, 107]. Although “benefits in terms of a rich picture of UX and higher scientific quality by collecting data with a combination of UX evaluation methods are well recognized” [107], there was a call for developed knowledge of which methods work well together and how to effectively collect and analyze the data from different sources.

Triangulation

As mentioned, the lack of multi-dimensionality in the evaluation formats was a commonly identified issue in previous meta-reviews. Law et al. [65] write “employing quantitative measures to the exclusion of qualitative accounts of user experiences, or vice versa, is too restrictive and may even lead to wrong implications”. Arhippainen et al. [8] demonstrate how applying several methods in practice can help researchers to learn about users and their ways to express experiences, and to catch “user experience information piece by piece by utilizing different methods”. However, they conclude that there is a general lack of knowledge generation in the area of UX and using multiple methods. Furthermore, it has been suggested that experience data is accessible in layers [108] and that therefore more of an experience can be understood by, for example, moving between “the expressible” by interviews [93], to tacit behaviors by observation techniques and latent experience data of “knowing, dreaming and feeling” by, for example, generative sessions [108].

Outside the UX field, the roots of method triangulation approaches traces back to the “paradigm wars” [104] and grew in popularity during the 1980s in social, behavioral and human sciences [44] to bridge different epistemological standpoints in research; applying triangulation approaches served as a way to overcome differences in approaches to knowledge production, enabling both the abilities of qualitative research in understanding the subjective, as well as the quantitative to determine statistical trends and connections. Denzin [19] outlined four types of triangulation in order to study a phenomenon: (a) data triangulation (i.e., the use of a mix of data sources in a study), (b) investigator triangulation (i.e., a number of researchers researching the same phenomenon), (c) theory triangulation

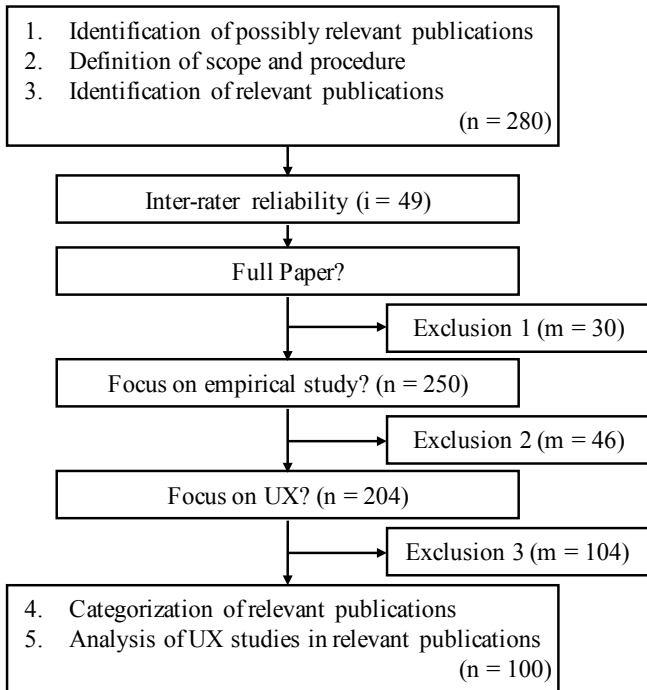


Figure 2. Procedure of our literature review, adapted from the QUOROM statement.

(i.e., the use of a number of theories used for interpreting results of a study), and (d) methodological triangulation (i.e., the use of more than one method to study a phenomenon). Creswell [18] describes the overarching two different types of employing two or more methods; either *sequential* (firstly either a quantitative or qualitative method is used, and the other type is used in a following study to explain, explore or validate the results) or *concurrent* (where two or more methods are employed within the same study to cross-validate findings). Employing triangulation to study a subject has been claimed to contribute to a more reliable, holistic and well-motivated understanding of phenomena [44], and to counteract inherent biases from data sources, investigators and especially methods.

Thus, triangulation can be claimed to lead to higher confidence in results and also to uncover unexpected results. This is true especially for multi-dimensional topics, such as UX, that may need multiple approaches, bringing depth as well as breath to the understanding. In their review of empirical studies, Bargas-Avila and Hornbæk found a “sad lack of reference” [9] between the groups of qualitative and quantitative methodological approaches. We are following up on the topic of triangulation approaches in UX evaluation - has there been progress? What can we learn from strategies currently applied?

STUDY AIM AND LITERATURE STUDY SETUP

UX evaluation continues to be a popular, but often debated, topic in conferences as well as in journals. Our goal is to provide an updated overview of UX evaluations in academia, from the years that have passed since the previous efforts of literature reviews, e.g., [9, 80, 90, 107]. We are interested in the methodologies employed, the dimensions of UX that are

Conference / Journal	Database (ACM DL)	Search Results (total)	Relevant Publications
CHI	7,482	137	50
UbiComp	2,265	34	10
DIS	1,092	40	14
CSCW	1,701	17	8
UIST	839	13	3
ICMI	843	1	0
IUI	837	20	7
TOCHI	278	18	8
<i>Total</i>	<i>15,337</i>	<i>280</i>	<i>100</i>

Table 1. Number of search results (excluding results from extended abstract, companion, and adjunct proceedings) and relevant publications for our analysis (per venue from 2010 to 2016).

studied, the products studied and the stages of development process in which the studies takes place. Furthermore, we investigated the application of triangulation within UX studies, whether it has increased since the turn of the decade, and identified current approaches for handling multidimensionality.

By evaluation, we mean a focus on assessing specific designs, from early concept ideas, over prototypes to finished products, in order to inform a design process. As our aim is to study the overall UX field, based on the academic literature, we did not select a specific practical model or view of UX as a basis for paper selection. Accordingly, we reviewed papers stating to evaluate UX as a part of their description of the study.

We decided to pursue a systematic analysis approach, based on a representative sample of publications in the field of Human-Computer Interaction and UX, to derive suitable insights regarding triangulation approaches in academic UX studies. Our approach (see Figure 2) is based on a procedure similar to Bargas-Avila & Hornbæk [9] and Lachner et al. [57] hence similarly adapted from the QUOROM statement [77], which specifies guiding principles on how to conduct meta-analyses including a quantitative data synthesis and a clarifying flow diagram.

Step 1: Identification of possibly relevant publications

Source selection. Academic work related to UX evaluation is spread across multiple scientific journals and conferences and continues to gain interest.

To limit the scope of our analysis we decided to only use the ACM Digital Library (DL) as a research database, including 476,316 records (207,571 from 2010 to 2016) in total at the time we conducted our review.

Furthermore, besides being a rich source of UX research, it contains influential conferences and highly ranked journals such as CHI, DIS and TOCHI. We further narrowed down the scope through our selection of suitable target conferences and journals. We based the selection of suitable venues on the h5-index indicated by Google Scholar as well as the relevance for our study aim hence omitted proceedings that target a specific domain (e.g., robotics, mobile, etc.). As a consequence, eight ACM conferences were selected (see Table 1).

Search procedure. We used the search query “*user experience AND “evaluation OR method OR measure OR assessment OR study”*” in any field and limited the search results to the period 2010 to 2016. We excluded the current year (2017),

since the publication year was not yet finished and thus not yet all possible relevant papers available. Next, we selected the conference proceedings (respectively the journal) of all selected venues and excluded all extended abstract, adjunct, and companion proceedings if they were listed individually (based on [9]). False positives were excluded at later stages. Our procedure led to 280 papers in total (see Table 1).

Step 2: Definition of scope and procedure

Exclusion criteria. As a next step, we defined criteria to exclude publications that were out of scope of our study aim. A publication was excluded if (1) it was not a full paper, or (2) if there was no trace of an empirical study, i.e., not including an empirical study of any kind, not including a clear description of the evaluation process or the study results, or (3) if the authors of the paper did not clearly state that evaluating UX (in any form) of a product or service was the aim of the empirical study. For the third exclusion criteria, it is important to note that we only included publications where the authors directly linked their study to UX while we excluded publications where the term “user experience” was only mentioned in the abstract, key words, related work, and/or introduction and not as a part of describing the specific study. We did not judge the authors’ views on UX; if authors claimed to study user experience, they were included in our detailed categorization.

Screening categories. Next, we selected the categories (based on [9]) that we used to analyze all relevant publications. Name and type of method were categorized (e.g., self-developed questionnaire, standardized questionnaire, free interview, semi-structured interview, activity tracking, live observation) and information if the method was referenced or not was noted. The (data) type of the methods in each publication (qualitative, quantitative, or both), task orientation (explorative use where the user was free to explore without guidance, or task-oriented), and if a motivation for the use of triangulation was stated (if two or more methods/types of data were applied). We also noted place of study (lab/field) and period of use (single-session/shortterm/longterm). For each study, we noted if there were references to UX literature and theory, as well as which dimension of UX that was studied. Thereby we differentiate between the consideration of following aspects: generic UX (experience is studied as an own construct without mentioning directly what is to be collected or measured), pragmatic quality (usability, functionality), hedonic quality (“psychological well-being through non-instrumental, self-oriented product qualities” [20, 36]), aesthetics/appeal, satisfaction, affect/emotions, enjoyment/fun engagement/flow, frustration, motivation and other constructs. We elaborate further on the categories in the results section.

Procedure. A common understanding and acceptable inter-rater reliability of the definition and interpretation of the exclusion criteria and the screening categories was ensured by four cross-checks before the final screening, containing of in total 60 papers mutually reviewed in full. Each of the the four cross-check rounds consisted of an independent analysis as well as a joint telephone conference of the first three authors to discuss 15 papers from each of the venues CHI ’16, CHI ’15, DIS ’16, and DIS ’14. These cross-check helped to decide how

to interpret the inclusion or exclusion of papers and how to assign the defined categories. Most prominently, we sharpened selection criteria 3 and decided to exclude evaluations that mainly target the analysis of experiences with technology in general, e.g., with the aim to derive UX theory, as compared to the evaluation of a specific product or product type. Whereas Hayashi and Hong [41], for example, state that “the overall goal of the studies was to investigate the user experience in using an authenticator [...]” (included in our review), Tuch et al. [101] start their survey study with the question “Bring to mind a single outstanding positive experience you have had recently” (omitted). Similarly to the latter, Mekler & Hornbaek [72] aim “to identify hedonic and eudaimonic components of [...] experiences”. Both latter examples were excluded from our review for stringency to our meta-study by a clear focus on the directed evaluation/exploration of a product/service type or case.

Step 3: Identification of relevant publications

Inter-rater reliability. We considered the exclusion of publications as crucial for the subsequent analysis. To ensure the reliability of the selection process, the first three authors conducted a final screening test round. In the test round, each author individually screened the same set of 49 papers (17,5% of all possibly relevant publications). The set consisted of randomly selected papers of each conference/journal and year. The inter-rater reliability for the exclusion was found to be $\alpha = .8307$, 95% in a CI of (0.7161, 0.9345). According to Krippendorff [53], values for α higher than .8 can be seen as a satisfactory.

Procedure. For the final screening procedure, we split all publications of papers between the first three authors with weekly meetings to discuss borderline papers.

Exclusion. Firstly, 30 publications were excluded because they were not full papers. Second, 46 papers were excluded because they were not empirical studies of a specific product or product type. Third, 102 papers were excluded for not relating the concept of UX to the empirical study (see Figure 2).

Step 4: Categorization of relevant publications

After identifying all relevant publications, the three authors categorized the same set of publications according to the defined screening categories. Once again, weekly meetings were held to handle unsure cases of categorizations.

Step 5: Analysis of UX studies in relevant publications

Interim analysis. After about half of the time needed for the categorization in step 4, we organized a workshop at DIS 2017 [84] to discuss initial insights with researchers in the UX field. At the workshop, we presented first insights of our review, including, for example, types of products studied, UX dimensions addressed, referenced UX theory, employed methods, and triangulation approaches. Together with all workshop participants, we interpreted our initial findings at that time. These interpretations provided an initial basis for our final screening and analysis process.

Final analysis. Finally, we finished screening and categorizing all relevant publications. The final screening also included

one more cross-check for borderline papers. For the analysis, we first looked at general developments in the field of UX evaluation. Second, we specifically examined mixed method approaches/triangulation patterns.

RESULTS

Below we report our results in following structure: We begin by describing general insights about the development of the amount of UX publications over time (1), the studied dimensions (2), and context (3). Then, we present our analysis of the applied methodology (4), which ends up in detailed insights about common triangulation patterns (5) in UX studies.

The Development of User Experience

In general, we see a temporal development of the papers we rated as relevant (empirical studies with a focus on UX) with a percentage increase of 283% from 2010 to 2016. This means that in total 40% (N=100) (of all identified full papers (N=250)) were identified as relevant for a detailed categorization. However, the percentage of all full papers that have been excluded because they did not focus on UX (in total 41%) converges over the years with the percentage of papers we rated as relevant. Looking at the linear trend lines, there is only a slight decrease (-2.5%) of papers rated as relevant, and a slight increase of “no UX papers” (5.1%). Thus, neither a positive nor a negative development was observed over time (in terms of amount of relevant papers, i.e., the numbers of publications containing the search words increased continuously) but the percentage of papers found relevant in relation to the total numbers of papers containing the search words, remained fairly stable. This means that there is a continuous growth of UX studies. Below, we will discuss some insights regarding the UX dimensions we studied, the context and the methodology we used.

UX Dimensions

UX is mostly studied as a general construct

In the detailed analysis and categorization of relevant papers (N=100), we found that the category that we summarized as *Generic UX* was the most frequently evaluated UX dimension (56%, see Table 2 for a full list of all UX constructs we used). In the *Generic UX* category, UX authors understood UX as a general construct and did not specify which aspects they studied in detail. For example, Woo et al. [111] describes: “we conducted a qualitative user study to understand people’s

experiences with DIY smart home products”. In Bargas-Avila and Hornbæk’s study of papers from 2005-2009 [9], *Generic UX* was also the main experiential dimension, yet slightly less prominent (in 41% of all papers). Consequently, there has been an increase from 2010 to 2016 of papers that evaluate UX as a broader construct. In 2010, only 1 out of 6 papers studied UX as a general construct [49]. In 2016, 52% of the papers did not define any additional concrete dimensions.

In addition, 22% of all relevant papers measured the pragmatic quality, by constructs of usability, ease of use, and/or efficiency. The reasoning behind focusing on the pragmatic quality differed, e.g., “to better understand the impact on the User Experience, we conducted a lab-based user study to evaluate the effectiveness of different time series visualizations that use varied interaction techniques, visual encodings and coordinate systems for four tasks [...]” [1]. Other constructs which were understood as a dimension of UX included Aesthetics/Appeal (7%), hedonic quality (6%) and satisfaction (5%). Enjoyment/Fun and Affect/Emotion, both considered as core dimensions in [9], are only investigated in 4% of all papers. Engagement/Flow, Frustration and Motivation were also rarely studied.

Rare links to UX theory

The high percentage of papers using a vague description of UX is also reflected in the theoretical frameworks of the papers. Overall, only 17% of all papers use a definition of UX, 83% do not. Furthermore, established UX theory papers, e.g., [21, 23, 64, 65, 37, 38] are only referenced extensively in 8% of all papers, to some extent in 17% and in 75% not at all. Häkkinen et al. [32] wrote “[...] Although there is hardly a unified definition for UX [62, 64], it is widely agreed that UX goes beyond usability and instrumental aspects [64]. A definition presented in [38] describes UX as ‘a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service’[...]”, as an example of a paper containing extensive references to UX theory. Such extensive descriptions were rarely found in all relevant papers.

Context

Wide range of products studied

While in 2005 to 2009, art was the most frequently evaluated product [9], we found a very wide range of products studied from 2010 to 2016. 20% of all papers evaluated individual products - too diverse to create their own categories, e.g., icons [96] or interactive museum installations [42]. The most frequently studied product types were mobile phone/app (15%), followed by interactive games (13%), web tools (12%) and websites (10%). The UX of new technologies such as connected services/IoT (4%) and VR/AR (2%) are increasing, but the number of papers is still small, see Table 3.

Few studies focus on early product development stages

In total, 56% of all studied products are presented and discussed as prototypes (of which 96% are high-fidelity prototypes), 39% are finished products or beta versions. Earlier stages of concepts were rarely evaluated. Only two papers used a wizard-of-oz setting and only one paper analyzed stories in a narrated form.

UX dimensions	%**	Examples
Generic UX	56	[111],[75],[78]
Pragmatic Quality	22	[1], [109], [99]
Aesthetics/Appeal	7	[96]
Hedonic Quality	6	[99]
Satisfaction	4	[106]
Affect/Emotion	4	[54]
Enjoyment/Fun	4	[69]
Engagement/Flow	3	[4]
Frustration	2	[86]
Motivation	1	[100]
Other Constructs	16	e.g. trust [82]

Table 2. Dimensions of UX research. Note: *multiple dimensions in one study possible, based on all relevant papers (N=100).

Studied Products	%*	Examples
Mobilephone/App	15	[54], [67]
Interactive Game	13	[7], [3]
Webtool	12	[30], [83]
Website	10	[75], [100]
Professional Software	9	[99]
Audio/Video/TV	6	[32]
Connected Service/ IoT	4	[111]
Non-digital Product	3	[45]
Vehicle	3	[40]
VR/AR	3	[81]
Wearable	2	[24]
Other Products	20	e.g., icons [96]

Table 3. Products studied in UX research. Note: *multiple products in one study possible, based on all relevant papers (N=100).

Equal share of field and lab studies

In contrast to [9], where only 21% regarded the context of the product in their study, from 2010 to 2016, almost the half (45%) of the publications described a field study and thus involved the context in their investigations. Ghellal et al. [29], for example, studied the experience of an augmented reality game within a horror and vampire genre “merging a fictional universe and the physical environment into one pervasive experience, centering around a variety of augmented reality activities played out at sunset”. At the same time, 41% were lab studies. Remote studies were conducted by only 8% and a mixed setup (lab and field) by 4% (all percentages rounded). We found that professional software tools (77.8%; N=7), mobile phone/apps (53.3%; N=8) and connected services (75%; N=3) were mainly investigated in a field study, while interactive games (61%, N=8) are more frequently evaluated in a controlled lab setting.

Still limited “truly” longitudinal studies

Furthermore we can report that 63% of all selected UX studies are evaluating the UX within a single session. These sessions were mostly conducted in the lab (61.9%; N=39), or within a field study (25.4% N=16). However, at least 34% of all UX studies used a long term setup (several weeks). Of these, 76.4% are conducted at the field, 14.7% assess UX remotely and 5.9% are performed in a lab. Only 3% are analyzed in a short term setting (i.e., several days, thus longer than a single session but not for weeks or longer). In total we can speak of a positive development since 2010, where the studies stretching over several weeks were only available in “some papers” [9]. However, “truly” longitudinal studies which “cover typical product life cycles over several months and years” [9] are still missing at large, with important exceptions such as [55].

Method Application

Focus on traditional methods

In order to evaluate UX, a variety of methods from related fields, as well as newly developed methods have been employed over the years. With regards to method deployment in our data set (see Table 4), we observed that self-developed questionnaires were used in more than half of all papers (53%), 46% conducted semi-structured interviews, 31% employed activity logging, 26% used a standardized questionnaire, and 19% observed their users (see Figure 1). Probes (i.e., additional material given to the users to elicit experiences,

Method Type**	%*	Examples
Self-Developed Questionnaire	53	[81], [32], [75]
Semi-Structured Interviews	46	[24], [32],[111]
Activity Logging	31	[3], [99],[49]
Standardized Questionnaires	26	[40], [59]
Live User Observation	19	[96], [7]
Videorecording	16	[99], [45]
Free Interview	9	[102]
Think Aloud Feedback	6	[82]
Diaries	6	[7]
Focus Groups	5	[7]
Online Feedback	3	[66]
Probes	3	[79]
Physio-psychological	2	[25]
Others	5	e.g., sticky labels to capture context [68]

Table 4. Methods used in UX research. Note: *multiple methods in one study possible, based on all relevant papers (N=100) **data is collected without judging, whether the measurement is right for assessing UX.

such as the possibility to express experiences through video, photo or drawings) and objective measures such as physio-psychological methods were rarely used. Self-developed questionnaires are also the method which is most commonly used stand-alone (11%), followed by semi-structured interviews (9%) and standardized questionnaires (6%).

Broad range of specific methods

When looking at established methods that focus on a specific evaluation scenario, an analysis shows that there is a broad range; 40% use a unique method, which no other study employs in our dataset (see Table 5). The NASA-TLX questionnaire, developed to assess workload [35], was (perhaps surprisingly) the most frequently used method in all UX studies (7%). Next, the AttrakDiff questionnaire [39] (5%) and a second version of it (2%) were employed. The System Usability Scale was used in 3% of the papers. As a consequence, we cannot report a high consensus in methods in general.

Specific Method**	%*	Examples
NASA-TLX	7	[59]
AttrakDiff	5	[54]
System Usability Scale	3	[24]
AttrakDiff 2	2	[49]
User Engagement Scale	2	[4]
Aesthetics scale	2	[87]
PANAS	2	[52]
Others	40	e.g., SAM [76]

Table 5. Specific methods used in UX research. *multiple specific methods in one study possible, based on all relevant papers (N=100) **data is collected without judging, if the measurement is the right for assessing UX.

Data type is mostly mixed

Our systematical categorization shows that 32% of the data collected in UX studies is solely quantitative (e.g., activity logging, questionnaires, psycho-physiological data) and 22% solely qualitative (e.g., interviews, observations). However, combinations of different methods based on the same data type, meaning either two or more qualitative (respectively quantitative) methods, are part of these numbers. The bigger part of the papers (46%) applied both quantitative and qualitative measurements, meaning that the studies used two or more methods, i.e., a triangulation approach. This is further analyzed in the next section, both for method triangulation as well as also data triangulation.

Triangulation Patterns

Analyzing the methodology of the 100 selected papers, we can observe that the majority (72%) uses a triangulation approach. While 21% mix only the method (e.g., Campbell et al. [14] use activity logging during the interaction with weblog posts and a post-use self-developed questionnaire focusing on ease of use, enjoyment, and intention to return), in 46% also the data type (quantitative and qualitative methods) was triangulated. Only two papers mix theory [89] or user groups [83].

We were able to cluster and identify 8 insights about triangulation patterns based on our analysis of general method combinations (see Figure 2) and data type per temporal stage of assessment, i.e., before, during, or after the evaluated interaction (see Table 6) as described below.

Gaining deeper insights motivates triangulation

Our analysis shows that many authors justify the use of multiple methods. In 32 (44%) of all papers that use any kind of triangulation (N=72) the authors state a motivation for the use of multiple methods. Besides, e.g., Ardito et al. [7] who justify their approach based on related work in the field of triangulation, most authors briefly mention that their aim of applying more methods or data types was to get deeper insights (e.g., [17] or [105]). Additionally, a second main reasoning behind method triangulation was to better understand the results of other applied methods, e.g., using post-use interviews to make sense of observations (see [31]) or post-use interviews to make sense of video recordings (see [67]). A total of 23% of all papers reported only positive results from the study, while the rest reported mixed results or mainly negative outcomes of the evaluation. When including only papers that applied triangulation, 19% reported on only positive outcomes.

Tendency towards triangulating a few methods

Our analysis also shows that the majority of method combinations is only based on a small set of different methods (see Figure 1). More precisely, we saw that self-developed questionnaires are most frequently used together with activity logging (23%) or semi-structured interviews (20%). Furthermore, semi-structured interviews are often combined with activity logging (15%) or with standardized questionnaires (11%). A typical triangle of methods (not necessarily applied in isolation) is the combination of self-developed questionnaires, semi-structured interviews, and activity logging (10%). The most frequent triangulation pattern, which is used standalone (without any additional methods) is self-developed questionnaire and activity logging (9%), whereas 6% additionally use semi-structured interviews. Thus, compared to previous research [9], we see a substantial increase of the use of activity logging as a complement to more traditional self-reporting. However, when qualitatively reviewing the content of the method descriptions, results, analysis and discussions, we often found weak links between the conclusions drawn from both sources.

Preference for quantitative data

From all methodological approaches, post-assessment based on quantitative methods represents the most common approach (see Table 6). However, from 16 publications which use questionnaires after the experience, there was only one publication

No.	Comment	Temporal Stage			%*
		before	during	after	
1	only one stage / quantitative			■	16
2	during/after: quantitative		■	■	11
3	post-use triangulation			□■	9
4	only one stage / qualitative			□	7
5	during/after: mixed		□	□■	6
6	during/after: qualitative		□	□	6
7	all stages: qualitative	□	□	□	5
8	during/after: mixed		□■	□	5
9	only one stage / qualitative			□	4
10	during/after: mixed		□■	□■	4
11	all stages: quantitative	■	■	■	4
12	during/after: mixed		□■	■	3
13	during/after: mixed		□	■	3
14	during/after: mixed		■	□■	2
15	during/after: mixed		■	□	2
16	all stages: mixed	■	□	■	2
17	all stages: mixed	□	□■	□■	2
18	all stages: mixed	□■	■	□■	1
19	all stages: mixed	■	□■	□	1
20	all stages: mixed	■	■	□■	1
21	all stages: mixed	■	□■	□■	1
22	all stages: mixed	■	■	□	1
23	all stages: mixed	□	□	□■	1
24	before/after	■		■	1
25	before/after	■		□■	1
26	before/after	□		□■	1
Sum		22%	65%	96%	100

Table 6. Overview of the identified data type combinations in all selected publications based on the studied data type(s) per temporal stage of assessment (quantitative method(s) = ■, qualitative method(s) = □). Note: *combinations are numbered according to their frequency of occurrence, based on all relevant papers (N=100).

in which the authors combine standardized and self-developed questionnaires to evaluate UX. The other 15 papers use either self-developed or standardized questionnaires. Nevertheless, the triangulation of quantitative methods during and after the interaction represent also the second most common approach (see Table 6, no. 2). Furthermore, quantitative methods are often used in the test stage of a design process (in total 15 out of 27 publications that use approaches no. 1 or 2, during or after interaction, see Table 6, e.g., [40] or [11]) and for lab studies (in total 17 out of 27 publications that use approach no. 1 or 2, see Table 6, e.g., [2] or [3]).

Thus, as already observed in previous research [64] we can still see the strong links of many UX studies to momentary performance metrics and usability-style experiments.

Infrequent pre-/post-evaluation

Table 7 shows an emphasized interest in evaluations during and after the interaction. In particular, 96% of all studies assess UX after the product usage, whereas 65% of all studies assess the UX during the interaction. Although 22% of the relevant papers investigate UX before the actual use, only three publications from recent years pursue an expectation-focused approach focusing on the analysis of pre-use and post-use evaluation. Furthermore, 19% assess UX in all temporal stages. Uriu et al. [102], for example, conducted interviews before and after the assessed interaction plus video recording during and after the interaction. Their goal was to study the UX of a whole cooking support system. Current research emphasizes the focus on pre-use and post-use evaluation (e.g., [74]). Expectation is a key aspect of an experience, yet still rarely analyzed in academic studies.

Before	During	After	%	Examples
	♦	♦	42	[4], [7], [87]
	♦	♦	32	[92], [60], [88]
♦	♦	♦	19	[68], [100]
	♦		4	[27]
♦		♦	3	[66]
22%	65%	96%		

Table 7. Temporal aspects of UX evaluation: before, during and after interaction. Note: *combinations are sorted by frequency of occurrence, based on all relevant papers (N=100).

Interviews and questionnaires for post-use evaluation

Interviews and questionnaires are not only two of the most common method types that were used in all relevant publications, but also the preferred triangulated methods, as previously stated. In our analysis, we see a tendency towards post-use triangulation of interviews and questionnaires, either as a stand-alone data type triangulation (see no. 3 in Table 6, e.g., [34] or [12]) or in combination with the additional evaluation before and/or during the interaction. In total, one quarter of all analyzed publications triangulate only questionnaires and interviews to evaluate the experience afterwards.

Vermeeren et al. [107] also observed that scale-based questionnaires, often have a follow-up interview to better understand research findings. In contrast, Alves et al. [5] more recently outlined that in practice, companies prefer observation and think-aloud over questionnaires and interviews.

Tendency towards more methods for exploration

It would seem plausible that the more data we collect, the more insights we can derive. Vermeeren et al. [107] question why researchers always want more data and suggest to rather focus on suitable combinations of methods. Our analysis confirmed the tendency towards applying more methods in UX studies, as only 28% of all relevant papers base their user study on only one method. Furthermore, the studies that pursued an explorative approach (i.e., with the main goal to explore a product or prototype freely rather than evaluating a specific task) tend to be based on more methods than task-oriented studies. From all 16 publications that use 4 or more methods in their empirical study, 12 (75%) pursue an exploratory user study (e.g., [17] or [46]). When we had a closer look at the papers that only applied 1 method in their study, we saw that 11 (39%) out of 28 publications focus on Generic UX (e.g., [2] or [81]). We agree with Vermeeren et al. [107] that it is not a cause in it self to add more methods, but careful consideration of combining the right method is key. We, however, consider the tendency towards more methods to be a generally positive trend in the exploratory studies to better understand the results.

Extensive long term studies and formative evaluation

The joint evaluation of expectations, UX during the interaction, and post-use UX is usually time-consuming and costly, but pointed out as an important key understanding of user experiences [47, 70]. Based on our analysis, we can see that such holistic evaluations of UX before, during, and after the interaction are often conducted for long term studies. From all relevant publications in our analysis, 19 analyze all temporal stages whereof 13 (68%) focus on a long term evaluation.

We had assumed to find more holistic studies addressing all temporal stages of UX evaluation, since their importance and value have been highlighted before [48]. However, similar to our findings, the review of Bargas-Avila and Hornbæk [9] in 2010 highlighted 17% of papers which analyzed all temporal stages and 3% that focused their evaluation on pre-use and post-use experience.

Diverse approaches for evaluations of accumulative UX

Almost one fifth (19%) of all relevant papers analyze all temporal stages, including the experience before the interaction (i.e., expectations), during the actual use, and post-use UX. Although about half of the publications that analyze all temporal stages focus either only on quantitative methods or only on qualitative methods (see Table 6), we identified a variety of different data type triangulation approaches. While Shin et al. [95], for example, base their study on mixed data pre-use, quantitative during use and mixed data post-use, Park et al. [82] use quantitative methods before and after the interaction and qualitative methods during the interaction. Further combinations are summarized in Table 6.

DISCUSSION

During the process of writing this paper, we presented our results at a workshop [84] and discussed them with UX experts of academia and industry (N=8). This helped us to critically analyze and assess existing approaches of UX evaluation method application from a practical and non-biased perspective.

How is the UX Research Field Evolving?

While the overall number of UX studies is increasing, only a quarter of the papers make any kind of reference to UX-specific literature. This raises the following question: Is the theory of UX already taken for granted or is it too vague or unknown? We had to exclude 104 papers which used the term “User Experience” as a buzzword but did not address the topic at all (based on our understanding). Often the term was even used within the title or the author keywords but nowhere else in the paper. One of the workshop participants stated: “UX is gaining attention, everybody wants to say they do UX – even though they don’t do it”. The fact that theory was often only vaguely addressed likely had consequences on the quality of evaluation. UX evaluation still appears to struggle with the same issues as in previous meta-reviews (e.g., lack of theory, lack of validated methods, overlap with usability), and one could question whether the field is maturing or disintegrating.

Furthermore, the lack of reuse of UX-specific methods was apparent in our data. The reasons behind this could not be revealed in our empirical data, but perhaps the nature of specific experiences (e.g., of a mobile health service [10] or a navigation system [43]) may not appear to be translatable to more generic methods for approaching the evaluation topic, and the researchers turn to, for example, self-developed questionnaires rather than reusing existing, validated ones. UX is a diverse topic and hence it may be misleading to look for a “one solves it all” method, but rather choose more specific methods for the specific type of experience and/or triangulation to accomplish a useful UX evaluation.

How Can we Exploit Data-Driven Methods for UX?

Surprisingly, the data-driven and objective method “activity logging” belongs to the 4 most frequently used methods in UX studies. But is it really a valid measurement to assess users’ experiences? This was also discussed at the workshop, and one participant stated: “*Activity logging is only used to report data, but relationships are rarely investigated*”. Similarly, we found a large number of papers employing user observations, but often it was not clear how observations were analyzed and how they actually contributed to the results. Exceptions were, of course, found, which contributed to a better understanding of the experience (see, for example, [94]). We believe that there is much more work to be done in this area, to guide technological efforts that can be of value when studying UX. Dove et al. [22] write: “*It is no longer enough for UX designers to only improve user experience by paying attention to usability, utility, and interaction aesthetics.*” and suggest that much more can be done to improve UX by employing machine learning for offering new value, such as personalization of systems by learning from user interactions. We found very little efforts in this area in relation to evaluation, and look forward to further progress of data-driven methods to help us understand UX, finding patterns and relationships instead of isolated data. We also found very few studies addressing different other emerging technologies and contexts of evaluation, such as UX evaluation by, for example, AR or VR, although it appears to be on the rise (see for example [28, 33] for recent studies), as well as the UX of interacting with virtual actors (see [15, 110] for examples) and automated systems [85, 73, 26].

For data-driven methods, we believe that triangulation of qualitative insights and quantifiable measures is important, either by sequential triangulation (i.e., using qualitative data to understand identified patterns in quantitative data or the other way around [18]) or concurrent triangulation, to better grasp and validate the UX data as it is gathered.

Why are Early UX Evaluations still rare?

Michalco et al. [74] and Kujala et al. [55] suggest to study the relation between expectations and UX as expectation disconfirmations, which has a significant effect on the overall UX. However, our study revealed that as of today, only few publications investigate the relation between expectations and the post-use experience (assessment before and after the experience). The infrequent comparison between expectations and after-use evaluation is also related to the persisting lack of UX evaluation at early design process stages, although this is often claimed to be important in a design process [90, 107, 9]. We had assumed to find more of these studies, but we did not, even if the examples of methodological approaches of Wizard of Oz (i.e., a human controlling the interface to respond to a user in a test setting) used in the field [17, 31] were very informative.

We conclude that most studies are focused on tangible and complete or almost complete designs; early stage evaluation relies heavily on the imagination of the study participant and is a step researchers may be unwilling to take. In addition to this, we were interested to see if there were approaches uncovering

not only the expressible and readily available responses to an experience, but also the tacit and the latent aspects of experiences (c.f. [93]). Of all papers, 4% used extra stimuli/probes during the evaluation, such as the possibility to express the experience in video and audio material (see [68]). We think that these are interesting approaches deserving further exploration.

Parallel Analysis rather than Triangulation?

Our analysis of sequential and concurrent triangulation patterns (cf. Creswell, [18]) demonstrated the complexity and variety of UX evaluation. In this area we could see most progress in the research field, but also identified several methodological gaps. In the papers we reviewed, there were examples of well executed sequential triangulation. Sequential triangulation holds the possibility of ensuring a systematically user experience-driven process, where initial findings can be followed up by additional data for further explanation, validation or exploration [18]. As examples of studies where data was explored sequentially, Leong et al. [68] used an initial diary study as later basis for further explanation during interviews with the participants. Hart et al [34] challenged their results from questionnaires based on unexpected findings in qualitative interview data. Hayashi and Hong [41] validated their primary data source by deriving from the quantitative data that the participants had a reasonable amount of exposure of the system to evaluate it qualitatively. Results like these serve an important role in building confidence for the data validity. Triangulation of sequential exploration of themes found in initial user studies were common, such as Lederman et al. [66] who employed sequential exploration by first understanding a product space qualitatively, triangulated with the evaluation of a designed prototype. This was a more commonly applied type of triangulation, than for example validation of data points across data sources.

We found very few examples of concurrent triangulation that carefully matches quantitative with qualitative data to derive a truly joint analysis, where results can be questioned or strengthened based on correlations or the lack thereof in the data. In many papers reviewed that applied triangulation, different types of data are gathered, but rarely cross-analyzed. Many studies left us with questions whether there was not more to be learned from the data with regards to correlation or differences between different types of data. There appears to be a growing understanding in UX research that using more than one method is beneficial, but a well-grounded knowledge about how to systematically cross-analyze data appears less widespread. Actively analyzing overlaps and differences in, for example, qualitative and quantitative data, that can add to richer and better validated knowledge of the evaluated topic, is still rare. Valuable exceptions are however for example Woo et al. [111] who make connections between data points over time as well as from different formats, to strengthen the outcomes.

Many studies provide a very short and general motivation of triangulation, if any, but there are exceptions, such as Kim et al. [51] who describe the process of applying grounded theory for the understanding the nature of the experiences and numeric data of questionnaires and log data of system usage for providing descriptive statistics linked to the themes.

Ardito et al. [7] use a thorough motivation of their triangulation of data sources such as observations, questionnaires and focus groups, later cross-analyzing and making connections between the sources. Perhaps the “sad lack of reference” [10] between the quantitative and qualitative is beginning to lessen, but a widespread understanding how both approaches can contribute to each other apparently has not been accomplished yet. There could have been many more good examples of better integration of data during the analysis of results, as triangulation approaches have been claimed to lead to deeper understandings and sometimes unexpected results.

Outlook on UX Research

From our review, we conclude that there are a number of specifically open questions to be further addressed in the UX field: There is a need to provide **further guidelines and practical examples for effective combinations** of different methods, i.e., triangulation strategies in UX evaluation. These could further exemplify how results from studies can be explained either by the use of sequential triangulation, e.g., understanding a pattern found in quantitative data by employing further methods gathering qualitative data, or by concurrent triangulation, e.g., reinforcing findings with two or more types of data that are cross-analyzed. Along more theoretical descriptions, also practical examples of addressing UX from a triangulation perspective would be inspirational and serve as a palette of examples of methods to use.

More studies are required which analyze the relation between **expectations** and UX, building on the importance to satisfy user’s expectations to achieve high UX [55]. We believe that further incorporating expectations in UX evaluation studies is an interesting thread to follow up on in more evaluative papers. There is a lack of empirical papers addressing “multi touch-point” or “**multi-device**” experiences [91], although important for many products today [97], i.e., services that do not only link to one type of product and context of use. As this is an increasing part of our daily lives, it has gained interest in ACM publications (see for example [6, 16]). We also believe there is much work to be done to address the specific nature of multi-device UX. The topic also connects to multi-user environments, which were addressed in 16% of the studies. To encompass the needs of studying emerging experiences of, for example, the IoT, the palette of evaluation strategies must be expanded to encompass multi-device and multi-user experiences. The methods used for studying a single person or product in one context need to be challenged and expanded or accompanied by other approaches. Up to 2016, we found hardly any work concerning **up and coming technological approaches** such as machine learning, that, however, appears to be applicable in UX evaluation, given the possibilities of technological development [22, 98].

LIMITATIONS AND FUTURE WORK

In our review, sources outside ACM were excluded, which may have provided a bias towards approaches founded in engineering and human factors perspectives rather than, for example, a design, psychology or a marketing perspective on UX. Directing the search to, for example, more design-oriented conferences and journals will most likely provide

further insights into the state of UX evaluation. Further work could also encompass additional sources inside ACM that were now excluded, such as the NordiCHI conference series. Especially, there is a need to constantly keep the analysis up-to-date with each finished publication year, starting with 2017. UX is still a developing research field which needs to be observed continuously.

Given the substantial amount of research in the field, it would also make sense to narrow the focus of further meta-reviews. For example, it could be specific types of experience dimensions and types of data collected for these, enabling deep-dives into more isolated questions. An example of such a deep-dive is provided by Mekler et al. [71] on quantitative evaluations of enjoyment from interactive games. Even if more specificity is needed in meta-reviews, we also look forward to other reviews of where the field of UX is heading, not only in terms of academic evaluation but also, for example, in theoretical foundations and industrial practice (as only 16% of the reviewed papers had 1 or more authors with an industrial affiliation).

CONCLUSION

To analyze the current state of UX evaluation in academia, we systematically identified 280 relevant papers, out of which 100 papers were finally selected for full review. In the continuously growing number of papers over the years 2010-2016, an increasing diversity in this inherently multi-dimensional field can be found. This is, of course, an asset: a product may (at different stages of development) benefit from both “macro”- and “micro”-perspective evaluations [57]. However, we see that many of the challenges reported in earlier meta-reviews still remain, such as the weak links between theory and evaluation, little attention to expectations in UX, and a tendency towards self-defined questionnaires and post-use evaluation. Progress could be identified in the use of triangulation, by inclusion of more methods as well as a larger number of studies performed in field contexts. Thus, we can perhaps see land on the horizon, but conclude that UX evaluation currently still remains sort of a Bermuda Triangle, often depending on personal perceptions of UX rather than aggregated theory.

We see that interest and efforts in the UX field still persist, and we look forward to further work. Areas which need to be addressed are evaluation approaches to multi-device experiences, machine learning, upcoming technology for virtual experiences, and addressing expectations in UX. Method triangulation needs to be used more coherently; for stronger results in UX studies, we recommend to integrate and structure data better. For example, a well-defined structuring of observational data, improved cross-analysis of qualitative and quantitative data, and a solid definition, which aspect of UX to evaluate, will bring results forward.

REFERENCES

1. *Muhammad Adnan, Mike Just, and Lynne Baillie. 2016. Investigating time series visualisations to improve the user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5444–5455.

2. *Sarah Fdili Alaoui, Baptiste Caramiaux, Marcos Serrano, and Frédéric Bevilacqua. 2012. Movement qualities as interaction modality. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 761–769.
3. *Florian Alt, Stefan Schneegass, Jonas Auda, Rufat Rzayev, and Nora Broy. 2014. Using eye-tracking to support interaction with layered 3D interfaces on stereoscopic displays. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 267–272.
4. *David Altimira, Florian Floyd Mueller, Jenny Clarke, Gun Lee, Mark Billingham, and Christoph Bartneck. 2016. Digitally augmenting sports: An opportunity for exploring and understanding novel balancing techniques. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1681–1691.
5. Rui Alves, Pedro Valente, and Nuno Jardim Nunes. 2014. The state of user experience evaluation practice. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 93–102.
6. Craig Anslow, Pedro Campos, Andrés Lucero, Laurent Grisoni, Mirjam Augstein, and James Wallace. 2016. Collaboration Meets Interactive Surfaces and Spaces (CMIS): Walls, Tables, Mobiles, and Wearables. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces (ISS '16)*. ACM, New York, NY, USA, 505–508.
7. Carmelo Ardito, Maria F. Costabile, Antonella De Angeli, and Rosa Lanzilotti. 2012. Enriching Archaeological Parks with Contextual Sounds and Mobile Technology. *ACM Transactions on Computer-Human Interaction* 19, 4 (2012), 1–30.
8. Leena Arhipainen, Minna Pakanen, and Seamus Hickey. 2013. Mixed UX Methods Can Help to Achieve Triumphs. In *Proceedings of CHI 2013 Workshop 'Made for Sharing: HCI Stories for Transfer, Triumph and Tragedy'*. ACM, 83.
9. Javier A Bargas-Avila and Kasper Hornbæk. 2011. 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*. ACM, 2689–2698.
10. *Eric PS Baumer, Sherri Jean Katz, Jill E Freeman, Phil Adams, Amy L Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M Olson, and Geri K Gay. 2012. Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 475–484.
11. *Luca Benedetti, Holger Winnemöller, Massimiliano Corsini, and Roberto Scopigno. 2014. Painting with bob: Assisted creativity for novices. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 419–428.
12. *Gilbert Beyer, Florian Alt, Jörg Müller, Albrecht Schmidt, Karsten Isakovic, Stefan Klose, Manuel Schiewe, and Ivo Haulsen. 2011. Audience behavior around large interactive cylindrical screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1021–1030.
13. Mark Blythe, Marc Hassenzahl, Effie Law, and Arnold Vermeeren. 2007. An analysis framework for user experience (UX) studies: A green paper. *Towards a UX Manifesto* (2007), 1.
14. *Amy Campbell, Christopher Wienberg, and Andrew Gordon. 2012. Collecting relevance feedback on titles and photographs in weblog posts. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces - IUI '12*. ACM Press, New York, New York, USA, 139.
15. Heloisa Candello, Claudio Pinhanez, and Flavio Figueiredo. 2017. Typefaces and the Perception of Humanness in Natural Language Chatbots. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3476–3487.
16. Marta E. Cecchinato, Anna L. Cox, and Jon Bird. 2017. Always On(Line)?: User Experience of Smartwatches and Their Role Within Multi-Device Ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3557–3568.
17. *Sandy Claes, Karin Slegers, and Andrew Vande Moere. 2016. The Bicycle Barometer: Design and Evaluation of Cyclist-Specific Interaction for a Public Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5824–5835.
18. John W Creswell. 2013. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
19. Norman K Denzin. 1978. Triangulation: A case for methodological evaluation and combination. *Sociological methods* (1978), 339–357.
20. Sarah Diefenbach, Nina Kolb, and Marc Hassenzahl. 2014. The 'hedonic' in human-computer interaction: history, contributions, and future research directions. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 305–314.
21. ISO DIS. 2009. 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO)*. Switzerland (2009).
22. Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning As a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 278–288.

23. Jodi Forlizzi and Katja Battarbee. 2004. Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 261–268.
24. *Jutta Fortmann, Erika Root, Susanne Boll, and Wilko Heuten. 2016. Tangible Apps Bracelet: Designing Modular Wrist-Worn Digital Jewellery for Multiple Purposes. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, 841–852.
25. *Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. 2016. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2283–2294.
26. Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, and Clemens Schartmüller. 2017. Driving Hotzenplotz: A Hybrid Interface for Vehicle Control Aiming to Maximize Pleasure in Highway Driving. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 236–244.
27. *Chi-Wing Fu, Wooi-Boon Goh, and Junxiang Allen Ng. 2010. Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2213–2222.
28. Wei Gai, Chenglei Yang, Yulong Bian, Chia Shen, Xiangxu Meng, Lu Wang, Juan Liu, Mingda Dong, Chengjie Niu, and Cheng Lin. 2017. Supporting Easy Physical-to-Virtual Creation of Mobile VR Maze Games: A New Genre. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5016–5028.
29. *Sabiha Ghellal, Ann Morrison, Marc Hassenzahl, and Benjamin Schaufler. 2014. The remediation of nosferatu: exploring transmedia experiences. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 617–626.
30. *Nitesh Goyal and Susan R Fussell. 2016. Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 288–302.
31. *Erik Grönvall, Sofie Kinch, Marianne Graves Petersen, and Majken K Rasmussen. 2014. Causing commotion with a shape-changing bench: experiencing shape-changing interfaces in use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2559–2568.
32. *Jonna R Häkkinen, Maaret Posti, Stefan Schneegass, Florian Alt, Kunter Gultekin, and Albrecht Schmidt. 2014. Let me catch this!: experiencing interactive 3D cinema through collecting content with a mobile phone. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1011–1020.
33. Daniel Harley, Aneesh P. Tarun, Daniel Germinario, and Ali Mazalek. Tangible VR: Diegetic Tangible Objects for Virtual Reality Narratives. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 1253–1263.
34. *Jennefer Hart, Alistair G Sutcliffe, and Antonella De Angeli. 2013. Love it or hate it!: interactivity and user types. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2059–2068.
35. Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50.
36. Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction* 19, 4 (2004), 319–349.
37. Marc Hassenzahl. 2005. The thing and I: understanding the relationship between user and product. *Funology* (2005), 31–42.
38. Marc Hassenzahl. 2008. User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine*. ACM, 11–15.
39. Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003*. Springer, 187–196.
40. *Renate Häuslschmid, Sven Osterwald, Marcus Lang, and Andreas Butz. 2015. Augmenting the Driver's View with Peripheral Information on a Windshield Display. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 311–321.
41. *Eiji Hayashi and Jason I Hong. 2015. Knock x knock: the design and evaluation of a unified authentication management system. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 379–389.
42. *Eva Hornecker. 2016. The To-and-Fro of Sense Making: Supporting Users' Active Indexing in Museums. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 2 (2016), 10.
43. *Shaohan Hu, Lu Su, Shen Li, Shiguang Wang, Chenji Pan, Siyu Gu, Md Tanvir Al Amin, Hengchang Liu, Suman Nath, Romit Roy Choudhury, and others. 2015. Experiences with eNav: A low-power vehicular navigation system. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 433–444.

44. R Burke Johnson, Anthony J Onwuegbuzie, and Lisa A Turner. 2007. Toward a definition of mixed methods research. *Journal of mixed methods research* 1, 2 (2007), 112–133.
45. *Jinyung Jung, Seok-Hyung Bae, and Myung-Suk Kim. 2013. Three case studies of UX with moving products. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 509–518.
46. Mohsen Kamalzadeh, Christoph Kralj, Torsten Möller, and Michael Sedlmair. 2016. TagFlip: Active Mobile Music Discovery with Social Tags. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 19–30.
47. Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean-Bernard Martens. 2009. User experience over time: an initial framework. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 729–738.
48. Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean Bernard Martens. 2010. Measuring the dynamics of remembered experience over time. *Interacting with Computers* 22, 5 (2010), 328–335.
49. *Tuula Kärkkäinen, Tuomas Vaittinen, and Kaisa Väänänen-Vainio-Mattila. 2010. I don't mind being logged, but want to remain in control: a field study of mobile activity and context logging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 163–172.
50. Joseph 'Jofish' Kaye. 2007. Evaluating experience-focused HCI. In *Proceedings of CHI 2007 Extended Abstracts on Human factors in computing systems*. ACM, 1661–1664.
51. *Tanyoung Kim, Hwajung Hong, and Brian Magerko. 2010. Design requirements for ambient display that supports sustainable lifestyle. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. ACM, 103–112.
52. *Martin Knobel, Marc Hassenzahl, Melanie Lamara, Tobias Sattler, Josef Schumann, Kai Eckoldt, and Andreas Butz. 2012. Clique Trip: feeling related in different cars. In *Proceedings of the designing interactive systems conference*. ACM, 29–37.
53. Klaus Krippendorff. 2004. *Content Analysis: An Introduction to Its Methodology* (second ed.). Sage.
54. *Sari Kujala and Talya Miron-Shatz. 2013. Emotions, experiences and usability in real-life mobile phone use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1061–1070.
55. Sari Kujala and Talya Miron-Shatz. 2015. The Evolving Role of Expectations in Long-term User Experience. In *Proceedings of the 19th International Academic Mindtrek Conference (AcademicMindTrek '15)*. ACM, New York, NY, USA, 167–174.
56. Sari Kujala, Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Evangelos Karapanos, and Arto Sinnelä. 2011. UX Curve: A method for evaluating long-term user experience. *Interacting with Computers* 23, 5 (2011), 473–483.
57. Florian Lachner, Philipp Nägelein, Robert Kowalski, Martin Spann, and Andreas Butz. 2016. Quantified UX : Towards a Common Organizational Understanding of User Experience. In *Proc. NordiCHI 2016*.
58. Carine Lallemand. 2015. *Towards consolidated methods for the design and evaluation of user experience*. Ph.D. Dissertation. University of Luxembourg, Luxembourg.
59. *Carine Lallemand and Guillaume Gronier. 2012. Enhancing User eXperience during waiting time in HCI: contributions of cognitive psychology. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 751–760.
60. *Airi Lampinen, Vilma Lehtinen, Coye Cheshire, and Emmi Suhonen. 2013. Indebtedness and reciprocity in local online exchange. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 661–672.
61. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 63–76.
62. Effie Law, Virpi Roto, Arnold POS Vermeeren, Joke Kort, and Marc Hassenzahl. 2008. Towards a shared definition of user experience. In *CHI'08 extended abstracts on Human factors in computing systems*. ACM, 2395–2398.
63. Effie Lai-Chong Law. 2011. The measurability and predictability of user experience. In *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 1–10.
64. Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold POS Vermeeren, and Joke Kort. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 719–728.
65. Effie Lai-Chong Law, Paul van Schaik, and Virpi Roto. 2014. Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies* 72, 6 (2014), 526–541.
66. *Reeva Lederman, Greg Wadley, John Gleeson, Sarah Bendall, and Mario Álvarez-Jiménez. 2014. Moderated online social therapy: Designing and evaluating technology for mental health. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 1 (2014), 5.
67. *Yong-Ki Lee, Youn-kyung Lim, and Kunpyo Lee. 2016. Timelessness. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16* (2016), 73–83.

68. Tuck W Leong, Frank Vetere, and Steve Howard. 2012. Experiencing coincidence during digital music listening. *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 1 (2012), 6.
69. *Craig M MacDonald and Michael E Atwood. 2014. What does it mean for a system to be useful?: An exploratory study of usefulness. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 885–894.
70. John McCarthy and Peter Wright. 2004. Technology as experience. *interactions* 11, 5 (2004), 42–43.
71. Elisa D Mekler, Julia Ayumi Bopp, Alexandre N Tuch, and Klaus Opwis. 2014. A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 927–936.
72. Elisa D Mekler and Kasper Hornbæk. 2016. Momentary Pleasure or Lasting Meaning?: Distinguishing Eudaimonic and Hedonic User Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4509–4520.
73. Thomas Meneweger. 2017. Experiencing Automation: UX and Practices with Automated Systems. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems (DIS '17 Companion)*. ACM, New York, NY, USA, 407–408.
74. Jaroslav Michalco, Jakob Grue Simonsen, and Kasper Hornbæk. 2015. An Exploration of the Relation between Expectations and User Experience. *International Journal of Human-Computer Interaction* 7318, July (2015), 150701125618002.
75. *Hannah J Miller, Shuo Chang, and Loren G Terveen. 2015. I LOVE THIS SITE! vs. It's a little girly: Perceptions of and Initial User Experience with Pinterest. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1728–1740.
76. *Pejman Mirza-Babaei, Lennart E Nacke, John Gregory, Nick Collins, and Geraldine Fitzpatrick. 2013. How does it play better?: exploring user testing and biometric storyboards in games user research. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1499–1508.
77. David Moher, Deborah J Cook, Susan Eastwood, Ingram Olkin, Drummond Rennie, and Donna F Stroup. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet* 354, 9193 (1999), 1896–1900.
78. *Timothy Neate, Michael Evans, and Matt Jones. 2016. Designing visual complexity for dual-screen media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 475–486.
79. *Marianna Obrist, Rob Comber, Sriram Subramanian, Betina Piqueras-Fiszman, Carlos Velasco, and Charles Spence. 2014. Temporal, affective, and embodied characteristics of taste experiences: A framework for design. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2853–2862.
80. Marianna Obrist, Virpi Roto, and Kaisa Väänänen-Vainio-Mattila. 2009. User experience evaluation: do you know which method to use?. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2763–2766.
81. *Thomas Olsson and Markus Salo. 2012. Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2779–2788.
82. *S. Joon Park, Craig M MacDonald, and Michael Khoo. 2012. Do you care if a computer says sorry?: user experience design through affective messages. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 731–740.
83. *Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 235–240.
84. Ingrid Pettersson, Anna-Katharina Frison, Florian Lachner, Andreas Rienner, and Jesper Nohhage. 2017. Triangulation in UX Studies: Learning from Experience. In *Proceedings of the 2016 ACM Conference Companion Publication on Designing Interactive Systems*. ACM, 341–344.
85. Ingrid Pettersson and Wendy Ju. 2017. Design Techniques for Exploring Automotive Interaction in the Drive towards Automation. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, 147–160.
86. *Erika Shehan Poole. 2012. Interacting with infrastructure: a case for breaching experiments in home computing research. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 759–768.
87. Katharina Reinecke and Abraham Bernstein. 2011. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 2 (2011), 8.
88. *Stefan Rennick-Egglestone, Sarah Knowles, Gill Toms, Penny Bee, Karina Lovell, and Peter Bower. 2016. Health Technologies' In the Wild': Experiences of Engagement with Computerised CBT. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2124–2135.

89. *Shyam Reyal, Shumin Zhai, and Per Ola Kristensson. 2015. Performance and user experience of touchscreen and gesture keyboards in a lab setting and in the wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 679–688.
90. Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Effie Law, and Arnold Vermeeren. 2009. User experience evaluation methods in product development (UXEM'09). In *IFIP Conference on Human-Computer Interaction*. Springer, 981–982.
91. Virpi Roto, Heli Väättäjä, Effie Law, and Rachel Powers. 2016. Experience Design for Multiple Customer Touchpoints. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 146, 3 pages.
92. *Yea-Kyung Row and Tek-Jin Nam. 2014. CAMY: applying a pet dog analogy to everyday ubicomp products. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 63–74.
93. Elizabeth B-N Sanders. 2002. From user-centered to participatory design approaches. *Design and the social sciences: Making connections* 1, 8 (2002).
94. *Magy Seif El-Nasr, Bardia Aghabeigi, David Milam, Mona Erfani, Beth Lameman, Hamid Maygoli, and Sang Mah. 2010. Understanding and Evaluating Cooperative Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 253–262.
95. *Jaemyung Shin, Bumsoo Kang, Taiwoo Park, Jina Huh, Jinhan Kim, and Junehwa Song. 2016. BeUpright: Posture Correction Using Relational Norm Intervention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 6040–6052.
96. *Johanna M Silvennoinen and Jussi PP Jokinen. 2016. Aesthetic Appeal and Visual Usability in Four Icon Design Eras. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4390–4400.
97. Adalberto L Simeone, Ming Ki Chong, Corina Sas, and Hans Gellersen. 2015. Select & Apply: understanding how users act upon objects across devices. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 881–896.
98. Samaneh Soleimani and Effie Lai-Chong Law. 2017. What Can Self-Reports and Acoustic Data Analyses on Emotions Tell Us?. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 489–501.
99. *Jacopo Staiano, María Menéndez, Alberto Battocchi, Antonella De Angeli, and Nicu Sebe. 2012. UX_Mate: from facial expressions to UX evaluation. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 741–750.
100. *S Shyam Sundar, Jeeyun Oh, Saraswathi Bellur, Haiyan Jia, and Hyang-Sook Kim. 2012. Interactivity as self-expression: a field experiment with customization and blogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 395–404.
101. Alexandre N Tuch, Rune Trusell, and Kasper Hornbæk. 2013. Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2079–2088.
102. *Daisuke Uriu, Mizuki Namai, Satoru Tokuhisa, Ryo Kashiwagi, Masahiko Inami, and Naohito Okude. 2012. Panavi: recipe medium with a sensors-embedded pan for domestic users to master professional culinary arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 129–138.
103. Kaisa Väänänen-Vainio-Mattila, Virpi Roto, and Marc Hassenzahl. 2008. Towards practical user experience evaluation methods. *Meaningful measures: Valid useful user experience measurement (VUUM)* (2008), 19–22.
104. Koen van Turnhout, Arthur Bennis, Sabine Craenmehr, Robert Holwerda, Marjolein Jacobs, Ralph Niels, Lambert Zaad, Stijn Hoppenbrouwers, Dick Lenior, and René Bakker. 2014. Design patterns for mixed-method research in HCI. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 361–370.
105. *Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. 2015. Sangeet swara: A community-moderated voice forum in rural india. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 417–426.
106. *Yolanda Vazquez-Alvarez, Matthew P Aylett, Stephen A Brewster, Rocio Von Jungendorf, and Antti Virolainen. 2016. Designing Interactions with Multilevel Auditory Displays in Mobile Audio-Augmented Reality. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 1 (2016), 3.
107. Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM, 521–530.
108. Froukje Sleeswijk Visser, Pieter Jan Stappers, Remko Van der Lugt, and Elizabeth BN Sanders. 2005. Contextmapping: experiences from practice. *CoDesign* 1, 2 (2005), 119–149.
109. *Qing Wang and Huiyou Chang. 2010. Multitasking bar: prototype and evaluation of introducing the task concept into a browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 103–112.

110. Rina R. Wehbe, Edward Lank, and Lennart E. Nacke. 2017. Left Them 4 Dead: Perception of Humans Versus Non-Player Character Teammates in Cooperative Gameplay. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 403–415.
111. *Jong-bum Woo and Youn-kyung Lim. 2015. User experience in do-it-yourself-style smart homes. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 779–790.

*Note: Papers which were analyzed in this paper and used as examples in the text are highlighted in the references with *.*
A full list of analyzed papers can be reviewed under: <https://uxtriangulation.wordpress.com/>