

## Research Article

# UX Evaluation with Standardized Questionnaires in Ubiquitous Computing and Ambient Intelligence: A Systematic Literature Review

Ignacio Díaz-Oreiro <sup>1,2</sup>, Gustavo López <sup>1,2</sup>, Luis Quesada <sup>2</sup>, and Luis A. Guerrero <sup>2</sup>

<sup>1</sup>Information and Communication Technologies Research Center, University of Costa Rica, San José 11501-2060, Costa Rica

<sup>2</sup>School of Computer Science and Informatics, University of Costa Rica, San José 11501-2060, Costa Rica

Correspondence should be addressed to Ignacio Díaz-Oreiro; [ignacio.diazoreiro@ucr.ac.cr](mailto:ignacio.diazoreiro@ucr.ac.cr)

Received 12 January 2021; Revised 26 February 2021; Accepted 17 April 2021; Published 4 May 2021

Academic Editor: Alice Chirico

Copyright © 2021 Ignacio Díaz-Oreiro et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Standardized questionnaires are well-known, reliable, and inexpensive instruments to evaluate user experience (UX). Although the structure, content, and application procedure of the three most recognized questionnaires (AttrakDiff, UEQ, and meCUE) are known, there is no systematic literature review (SLR) that classifies how these questionnaires have been used in primary studies reported academically. This SLR seeks to answer five research questions (RQs), starting with identifying the uses of each questionnaire over the years and by geographic region (RQ1) and the median number of participants per study (how many participants is considered enough when evaluating UX?) (RQ2). This work also aims to establish whether these questionnaires are combined with other evaluation instruments and with which complementary instruments are they used more frequently (RQ3). In addition, this review intends to determine how the three questionnaires have been applied in the fields of ubiquitous computing and ambient intelligence (RQ4) and also in studies that incorporate nontraditional interfaces, such as haptic, gesture, or speech interfaces, to name a few (RQ5). **Methods.** A systematic literature review was conducted starting from 946 studies retrieved from four digital databases. The main inclusion criteria being the study describes a primary study reported academically, where the standardized questionnaire is used as a UX evaluation instrument in its original and complete form. In the first phase, 189 studies were discarded by screening the title, abstract, and keyword list. In the second phase, 757 studies were full-text reviewed, and 209 were discarded due to the inclusion/exclusion criteria. The 548 resulting studies were analyzed in detail. **Results.** AttrakDiff is the questionnaire that counts the most uses since 2006, when the first studies appeared. However, since 2017, UEQ has far surpassed AttrakDiff in uses per year. The contribution of meCUE is still minimal. Europe is the region with the most extended use, followed by Asia. Within Europe, Germany greatly exceeds the rest of countries (RQ1). The median number of participants per study is 20, considering the aggregated data from the three questionnaires. However, this median rises to 30 participants in journal studies while it stays in 20 in conference studies (RQ2). Almost 4 in 10 studies apply the questionnaire as the only evaluation instrument. The remaining studies used between one and five complementary instruments, among which the System Usability Scale (SUS) stands out (RQ3). About 1 in 4 studies analyzed belong to ubiquitous computing and ambient intelligence fields, in which UEQ increases the percentage of uses when compared to its general percentage, particularly in topics such as IoT and wearable interfaces. However, AttrakDiff remains the predominant questionnaire for studies in smart cities and homes and in-vehicle information systems (RQ4). Around 1 in 3 studies include nontraditional interfaces, being virtual reality and gesture interfaces the most numerous. Percentages of UEQ and meCUE uses in these studies are higher than their respective global percentages, particularly in studies using virtual reality and eye tracking interfaces. AttrakDiff maintains its overall percentage in studies with tangible and gesture interfaces and exceeds it in studies with nontraditional visual interfaces, such as displays in windshields or motorcycle helmets (RQ5).

## 1. Introduction

User experience (UX) is currently a key factor in establishing the quality of a product or service [1–3]. Although UX could be described in different views and efforts have been made to propose a definition that unifies the different approaches [4–6], in this study, we will refer to UX as defined by ISO [7]: a person’s perceptions and responses resulting from the use and/or anticipated use of a product, system, or service. ISO definition includes users’ emotions, beliefs, physical, and psychological responses and considers UX also a consequence of brand image, presentation, system performance, the user’s internal and physical state resulting from prior experiences, attitudes, skills, and personality, among others.

To study UX, an essential element is the evaluation, which refers to the application of a set of methods and tools whose objective is to determine the perception about the use of a system or product. Methods can be classified according to three dimensions: qualitative vs. quantitative, attitudinal vs. behavioural, and context of use (natural or scripted use of the product, not using the product or hybrid) [8]. Within these methods, researchers rely on different instruments or tools to evaluate UX such as expert evaluation, observations, self-designed questionnaires, interviews, and standardized questionnaires to name a few of the most widely applied.

Standardized questionnaires are instruments applied within the quantitative and attitudinal methods. They are considered standardized, since they contain an invariable set of questions, exposed always in the same order, in which participants respond themselves to express their feelings and experiences regarding different aspects of a product [9]. This makes them inexpensive and easy to use, since they are self-applied by the user based on the perceived or anticipated experience of using a product or service, and for this reason, its use is extended. In addition, they are considered reliable and valid to measure UX [10].

The three most recognized standardized questionnaires for UX evaluation are AttrakDiff, UEQ, and meCUE as it stated in studies such as those presented by Lallemand et al. [9, 11], Baumgartner et al. [12], Forster et al. [13], and Klammer and van den Anker [14].

These three questionnaires have been used in several primary studies reported in the academic literature, as can be seen in the 548 studies listed in this study, and within these, the use of this evaluation instrument in the fields of ubiquitous computing and ambient intelligence (as defined in Section 2) is of special interest to the authors of this study, as it is one of the study subjects of the research group we form. Studies in these fields will be categorized in this review in topics such as IoT and wearable sensors, smart cities, homes and other human-ambient interaction, intelligent transportation systems, indoor positioning and navigation, Internet of people, and smart environments for health, among others.

In a complementary way, the use of nontraditional interfaces (also defined in Section 2), such as haptic, gesture, eye tracking, small screen, voice, virtual reality, and robot interfaces, to name a few, is another research topic of our research group. In this context, when carrying out different

studies, the question always arises is if researchers around the world prefer a specific questionnaire over the others, if this preference is influenced by the topic under study or if there are trends among researchers when evaluating systems implemented with nontraditional interfaces. The purpose of this systematic review is to answer all these questions.

The studies mentioned above [11–14] do not address how the different standardized questionnaires have been used in primary studies, even though these studies cite the three questionnaires as the recognized scales for UX evaluation. These studies do separate UX from constructs such as usability, acceptance, and trust [12, 13] and describe the mechanics of use, application, and theoretical models on which the questionnaires are based [11, 14], but they do not provide details of the uses of the questionnaires. The only element that sheds some light on this subject is presented by Forster et al. [13] who conducted a search on Google Scholar and found 1157 citations of the three UX questionnaires. Of these citations, 697 correspond to AttrakDiff (60.24%), 429 to UEQ (37.08%), and 31 to meCUE (2.68%).

As for systematic literature reviews related to UX evaluation, we found two literature reviews on user experience evaluation in general, presented by Maia and Furtado [15] and Ten and Paz [16]. However, in neither of the two cases, authors established objectives as those formulated in our study. The review presented in [15] raises four research questions, of which number 2 (how is the evaluation performed?) could have been related to our SLR. Nevertheless, the authors focused on when to perform the evaluation, if it is performed manually or in an automated way and only mention the high use of questionnaires, but without mentioning the questionnaires used. In [16], a SLR is proposed to find the methods, tools, and criteria used to evaluate websites’ user experience. Although the study recognizes that questionnaires are the most used tool, it does not detail which questionnaire was used in studies. In fact, in none of these two reviews is there any mention of AttrakDiff, UEQ, or meCUE.

Due to this lack of information regarding the uses given to the different standardized questionnaires for UX evaluation, our systematic literature review is proposed. Since there is no general information on the uses of the questionnaires, there is even less data documenting the use of these standardized questionnaires in studies related to ubiquitous computing and ambient intelligence or in systems implemented using nontraditional interfaces, two topics of special interest to the authors of this study and which motivates the SLR presented in this study.

The following section describes the theoretical concepts used in this review. Section 3 describes the protocol used to carry out the systematic literature review. Section 4 shows the most important results of this investigation. Section 5 cites possible limitations of the research, and finally, and Section 6 presents the conclusions of this work.

## 2. Background

This section explains some characteristics of the standardized questionnaires AttrakDiff, UEQ, and meCUE (the

theoretical model on which each questionnaire is based, the structure and the subscales that make up each one) and also defines the concepts ubiquitous computing, ambient intelligence, and nontraditional interfaces as used in the context of this work.

Standardized questionnaires are widely used instruments to evaluate UX, composed of Likert scales [17] and semantic differentials [18]. The three most recognized standardized questionnaires in this field are AttrakDiff, UEQ, and meCUE [9, 11–14].

The first of the three questionnaires to appear was AttrakDiff, proposed by Hassenzahl, Burmester, and Koller in 2003 [19]. It consists of 28 items to be marked by the user, where each item is constructed by a 7-point semantic differential. Later, in 2008, Laugwitz, Held, and Schrepp presented the “User Experience Questionnaire” (UEQ) [20]. It consists of 26 items also built by 7-point semantic differentials. Finally, in 2013, Minge and Riedel proposed the meCUE questionnaire [21], built with 34 items: 33 7-point Likert scales and one 11-point semantic differential with the question “How do you experience the product as a whole?”.

The AttrakDiff questionnaire is based on the UX model proposed by Hassenzahl [22, 23]. It is composed of 28 items classified into four subscales: pragmatic quality, hedonic quality-stimulation, hedonic quality-identification, and attractiveness. Pragmatic characteristics refer to those traits as if a product is predictable, confusing, simple, and complicated, among others. On the other hand, hedonistic characteristics are those that appeal to feelings as if a product is boring, interesting, novel, or disappointing, related to stimulation traits and also to identification and evocation traits, such as the ability of a product to connect with others rather than to isolate [14]. Attractiveness describes the overall value of the product based on the perception of pragmatic and hedonic qualities [5].

The UEQ questionnaire is also based on the Hassenzahl model and consists of 26 items belonging to the subscales and factors such as attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty.

As both questionnaires are based on the same UX model, the subscales can be related as follows: perspicuity, efficiency, and dependability (UEQ) correspond with pragmatic quality (AttrakDiff), stimulation and novelty (UEQ) with hedonic quality-stimulation (AttrakDiff), and attractiveness (UEQ) with attractiveness (AttrakDiff). The AttrakDiff hedonic quality-identification scale would have no corresponding in UEQ [20].

For its part, the meCUE questionnaire is based on the Thüring and Mahlke [24] model. It is made up of 34 items corresponding to four modules, which in turn represent subconstructs: product perceptions (usefulness, usability, visual aesthetics, status, and commitment), user emotions (positive and negative), consequences of use (intention to use and product loyalty), and overall evaluation. Product perceptions refer to both instrumental perceptions (usefulness and usability) and noninstrumental perceptions (visual aesthetics, status, and commitment).

Ubiquitous computing (or UbiComp) refers to a genre of computing in which the computer permeates the life of

the user becoming a helpful but invisible force, assisting the user but without getting in the way [25]. The ubiquitous computing concept was proposed by Mark Weiser in 1991 to represent “a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background.” Weiser exposes the concept in the context of his work at the Xerox Palo Alto Research Center where he and his colleagues develop an office equipped with several ubiquitous computers of different sizes and with different tasks (called tabs, pads, and boards) that share information and communicate wirelessly with each other and with the people who use them. Weiser also presents some challenges to be solved in terms of software and communications required to connect the ubiquitous hardware. The goal is to have hundreds of computers per room, “machines that fit the human environment instead of forcing humans to enter theirs” [26].

The term ambient intelligence was coined by the European Commission ISTAG (Information Society Technologies Advanced Group) in 2001, with a vision of the information society in which “people are surrounded by intelligent intuitive interfaces that are embedded in all kinds of objects and an environment that is capable of recognizing and responding to the presence of different individuals in a seamless, unobtrusive, and often invisible way” [27].

These augmented spaces around the user can be open or a close environment, constrained in a physical location, or spread across a large space. As stated in [28], “the most important concept is that the pervasive network is able to track the user preferences through space and time, improving the human-machine relationship, constrained in a physical location, or spread across a large space.”

In [29], it is also highlighted that ambient intelligence represents technology that is “invisible, embedded in our natural surroundings, present whenever we need it, enabled by simple and effortless interactions, attuned to all our senses, adaptive to users and context-sensitive, and autonomous.”

Ambient intelligence concepts can be implemented in environments such as homes, hospitals, public transportation, education institutions, emergency services, and production-oriented places, workplaces, public spaces, playgrounds, and cities [30, 31].

Some authors consider that ubiquitous computing is one of the parts that make up ambient intelligence. For example, in [32], it is said that ubiquitous computing, ubiquitous communication, and intelligent user-friendly interfaces converge to form ambient intelligence. In [30], ambient intelligence is defined as composed of pervasive/ubiquitous computing, human-computer interfaces, artificial intelligence, sensors, and networks. On the other hand, [33] indicates that the terms ubiquitous computing and ambient intelligence are often considered interchangeable, where ubiquitous computing would be a more technical term, while ambient intelligence focuses “on the architecture and on more general aspects of how such vision could be integrated into human daily life.” In this work, both terms will be used together to encompass the concepts they cover,

without specifying in detail where the boundary between both concepts would be.

One of the distinctive factors in ubiquitous computing and ambient intelligence is the interaction with the environment through different interfaces present in all kinds of objects [27], which gives visibility and impulse to less common or traditional interfaces than GUIs (graphical user interfaces).

These so-called nontraditional interfaces can be categorized according to the number and diversity of inputs and outputs (communication channels). Each different independent channel is called a modality, and systems that are based on only one modality are called unimodal, while a system with multiple channels is called multimodal [34].

The nature of modalities divides them into visual-based, audio-based, and sensor-based. The following research areas cover visual-based modalities: facial expression analysis, body movement tracking, gesture recognition, and gaze detection. Audio-based modalities can be categorized into speech recognition, speaker recognition, auditory emotion analysis, human-made noise/sign detections (gasp, sigh, laugh, and cry), and musical interaction. Finally, in sensor-based modalities, there are pen-based interaction, mouse and keyboard (non-standard), joysticks, motion tracking sensors and digitizers, haptic sensors, pressure sensors, and taste/smell sensors [34].

Another categorization of nontraditional interfaces is presented in [35], where the following categories are proposed: haptic, gesture, locomotion, auditory, speech, interactive voice response, olfactory, taste, small screen, and two types of multimode interfaces. Two or more interfaces to accomplish the same task (mutually exclusive) and combining interfaces to accomplish a single task (mutually inclusive, like virtual reality interfaces).

### 3. Methods

The purpose of this systematic literature review is to collect information on the uses that have been given to the standardized UX evaluation questionnaires in academic studies with particular interest in the uses of the questionnaires in topics related to ubiquitous computing and ambient intelligence, as well as in nontraditional interfaces. We used the PRISMA statement for systematic reviews, as proposed by Liberati et al. [36].

*3.1. Planning the Review.* The objective of the following paragraphs is to document this SLR to make it replicable and auditable, so the research questions, the search strategy, and the studies inclusion and exclusion criteria will be presented next.

*3.1.1. Research Questions.* Based on the purpose defined above, the following five research questions (RQ) were defined:

RQ1. How have the standardized questionnaires AttrakDiff, UEQ, and meCUE been used to evaluate UX in primary studies reported academically?

RQ2. How many participants per study have been used in UX evaluations conducted with AttrakDiff, UEQ, and meCUE?

RQ3. What is the relationship of the standardized questionnaires with other UX assessment instruments applied in the same study?

RQ4. How have the standardized questionnaires been used to evaluate primary studies on ubiquitous computing and ambient intelligence?

RQ5. How have the standardized questionnaires been used to evaluate primary studies with nontraditional interfaces?

The first question (RQ1) is general in nature and aims to identify the number of times each standardized questionnaire has been applied, the uses per year, as well as per country and world region. In RQ2, we are interested in finding out the number of participants that have been used in UX evaluations and if there is a difference in studies published in journals from those published in conference proceedings. With RQ3, we seek to determine if the questionnaires are used as the only evaluation instrument or if they are applied in association with other instruments and if they are used in conjunction with other instruments, which of these would be the most frequently used.

In RQ4, we are interested in clarifying how the 3 standardized questionnaires have been applied in studies on ubiquitous computing and ambient intelligence to compare the preference of researchers in each questionnaire, in general and classifying the studies on different topics of ubiquitous computing and ambient intelligence. Finally, RQ5 aims to explore how AttrakDiff, UEQ, and meCUE have been applied in studies that use nontraditional interfaces if there are preferences on any of the questionnaires over the others, in general and categorizing the studies according to the different nontraditional interfaces.

*3.1.2. Search Strategy.* A preliminary study was carried out to establish the search strategy, to anticipate the number of articles that report uses of standardized questionnaires, as well as to organize the team that would oversee the revision. Additionally, this preliminary study allowed us to refine the search query, which was finally established as follows: (meCUE OR AttrakDiff OR AttrakDiff2 OR (UEQ AND (UX OR “User Experience”))). The acronym UEQ had to be associated with the concept UX or user experience to prevent the query from returning studies with matches for acronyms that did not correspond to the standardized questionnaire.

It is important to note that the query was constructed in such a way that it brought all the studies that mentioned any of the three AttrakDiff, UEQ, or meCUE questionnaires, regardless of the field of the study in which it was applied or the type of interfaces used. The selection of studies related to ubiquitous computing and ambient intelligence, as well as nontraditional interfaces, was carried out manually in the full-text review phase.



**3.1.3. Study Selection.** Before performing the search, the inclusion and exclusion criteria were defined as given in Table 1.

### 3.2. Conducting the Review

**3.2.1. Identification.** The search query was executed on ACM Digital Library, IEEE Xplore, Springer Link, and ScienceDirect, configured to run in the metadata as well as in the full text of the articles. A filter was added to the queries so that the answers did not include articles before 2003, the year on which AttrakDiff appeared, this being the first of the three questionnaires to be proposed. Additionally, for the Springer Link library, a second filter was set to left out articles in languages other than English. For other three libraries, discarding articles in other languages was performed manually during the screening process. As a result, 946 studies were collected, as can be seen in the upper level of the PRISMA diagram in Figure 1.

**3.2.2. Screening.** In a first phase, seven duplicate studies were eliminated as they were present in more than one digital library. Then, 182 studies were discarded based on the screening of the title, keyword list, and abstract, using the exclusion criteria given in Table 1. Even without reviewing the full text, it was possible to identify studies that did not meet the inclusion criteria, particularly studies where the full text was not available, the study proposed a new instrument or a variant of one of the three standardized questionnaires, or the study had nothing to do with UX evaluation, but the match was caused by coincidences with other UEQ and UX that are not the “User Experience Questionnaire” or the “User Experience” concept. Also, 10 studies written in languages other than English were discarded: 4 in French and 6 in Portuguese, of studies conducted in France and Brazil, respectively. Given that this amount is low in relation to the totality of the studies included, we do not consider that the exclusion of these 10 studies affects the results obtained in the work.

This screening process was carried out by four researchers, authors of this study. At the end of the process, a cross-checking process of the discarded studies was carried out by a different researcher from the one who had carried out the exclusion, and no discrepancies were found. As a result, a set of 757 studies was obtained to be reviewed in the eligibility phase.

**3.2.3. Eligibility.** As a result of the full-text review phase, 209 articles were removed, following the exclusion criteria indicated in Table 1, mostly as the study proposes a new instrument and uses the questionnaire as a basis or reference, the questionnaire used is a translation of one of the standardized questionnaires into another language, or the study mentioned one of the three questionnaires in the “related work” section, but the questionnaire is not used in the primary study. Nine studies were also excluded as they used more than one of the three questionnaires, either

because they are original studies presenting meCUE [37, 38], where they compare their results with AttrakDiff and UEQ, the original UEQ study [20], in which AttrakDiff is used for comparison purposes, or are studies comparing two of the questionnaires to explain concepts used in them [39–44]. The rest of the studies (548) did meet the inclusion criteria indicated in Table 1.

**3.2.4. Included.** Product of the eligibility phase, a total of 548 studies were included in the qualitative analysis. Table 2 provides a summary of the studies reviewed and included for each phase of the process, classified by the digital library to which they correspond.

The research was conducted mostly by a group formed by the four authors of this study, reviewing 135 studies each, on average, and to a lesser extent by a second group formed by master’s degree students, carrying out the complete text review of 20 articles each student. It is worth noting that cross-review processes were performed by different researchers, from the first group, in all phases, and discrepancies were settled.

It should be mentioned that four studies presented more than one study, so the total number of studies is 553. The researchers analyzed these 553 studies in detail and obtained the results presented in the following section.

## 4. Results and Discussion

This section presents the results obtained in the full-text study of the included articles, organized according to the information analyzed for each research question, following the order in which the questions were defined in Section 3.1. Initially, Section 4.1 shows three results of a general nature: number of questionnaires uses, trends in use by year, and geographic distribution of use. Next, Section 4.2 presents the results of the number of participants per study. In Section 4.3, the association of questionnaires with other evaluation mechanisms is shown. Then, Section 4.4 presents the results covering the use of standardized questionnaires on topics related to ubiquitous computing (UbiComp) and ambient intelligence. Finally, Section 4.5 shows the results related to studies that are implemented through nontraditional interfaces and to studies that cover both elements: UbiComp and ambient intelligence topic and nontraditional interfaces.

**4.1. General Characteristics of Use of Standardized Questionnaires (RQ1).** As given in Table 3, the AttrakDiff questionnaire was the most present questionnaire in the literature, being used in 341 of the 553 studies analyzed (61.6%). It is followed by UEQ with 200 studies (36.2%) and finally meCUE with 12 (2.2%).

Considering the year in which each questionnaire was presented (AttrakDiff in 2003, UEQ in 2008, and meCUE in 2013), one might initially think that the seniority of the questionnaire affects the number of uses reported.

If the progression of the total uses of questionnaires is reviewed, Figure 2 shows that it has been increasing through the years. This figure presents the data up to 2018, given that

TABLE 1: Inclusion and exclusion criteria.

	Criteria
Inclusion	The study describes a primary academic study where the standardized questionnaire is used as a UX evaluation instrument. The study is written in English, and the full text is available. The study uses the standardized questionnaire in its original and complete form.
Exclusion	The study is written in languages other than English. Full text is not available. The study is dated before 2003. Only mentions one of the standardized questionnaires, but it is not actually used in the study. The study proposes a new method and uses the questionnaire as a basis or reference. The study presented uses just a part of the standardized questionnaire (some of the items) for the evaluation. The study proposes a questionnaire that is a translation of one of the standardized questionnaires into another language.

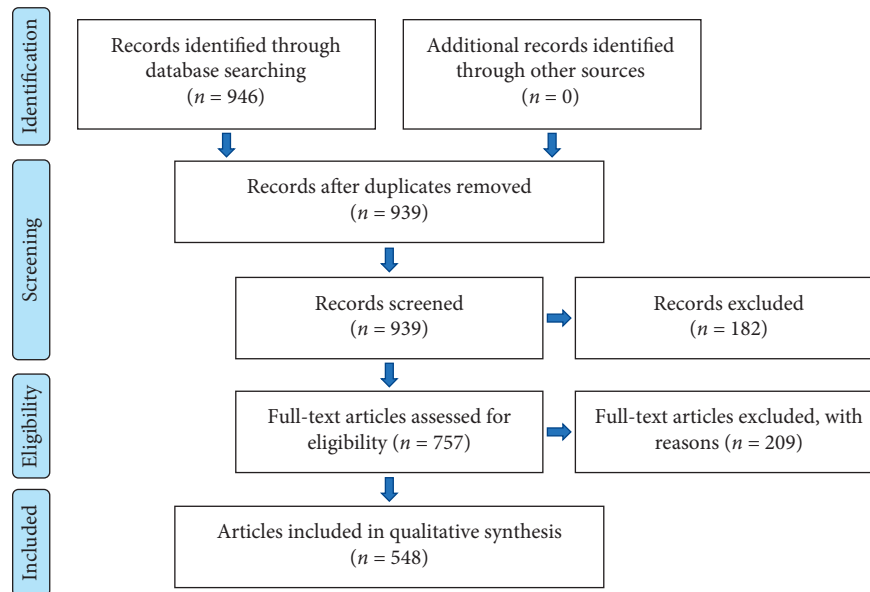


FIGURE 1: PRISMA flow diagram for this systematic literature review.

TABLE 2: Distribution of studies by source.

Source	Query result	Title and abstract reviewed	Title and abstract included	Full study reviewed	Full study included
ACM	351	348	304	304	234
IEEE Xplore	145	144	132	132	94
Springer link	343	340	250	250	166
Science direct	107	107	71	71	54
Total	946	939	757	757	548
Percentage of total	100.0%	99.3%	80.0%	80.0%	57.9%

TABLE 3: Total number of uses of standardized questionnaires.

Questionnaire	AttrakDiff	UEQ	meCUE	Total
Uses	341	200	12	553
Percentage	61.6%	36.2%	2.2%	100%

the consultation of the different libraries was performed in March 2019, so the results for 2019 did not encompass the full year. It is also important to note that, although AttrakDiff was presented in 2003, it was not until 2006 that the first studies using it were published.

Analyzing the number of uses of each questionnaire individually, it can be noted that AttrakDiff, being the first

questionnaire to appear and that globally covers 62% of the uses, was surpassed in 2017 and 2018 by UEQ. While AttrakDiff has maintained a stable number of uses since 2015, UEQ is growing at a faster pace, surpassing AttrakDiff in 2017 and 2018 by 42% and 47% of uses, respectively.

Regarding meCUE, which appeared in 2013, it shows an increase in use in 2017 and 2018. As the number of studies using meCUE is still low, the behavior of this trend should be followed in the coming years to see if it manages to take a significant place next to AttrakDiff and UEQ, which are the most used to date.

As for the geographical distribution of use of the questionnaires, Table 4 provides that Europe is by far the

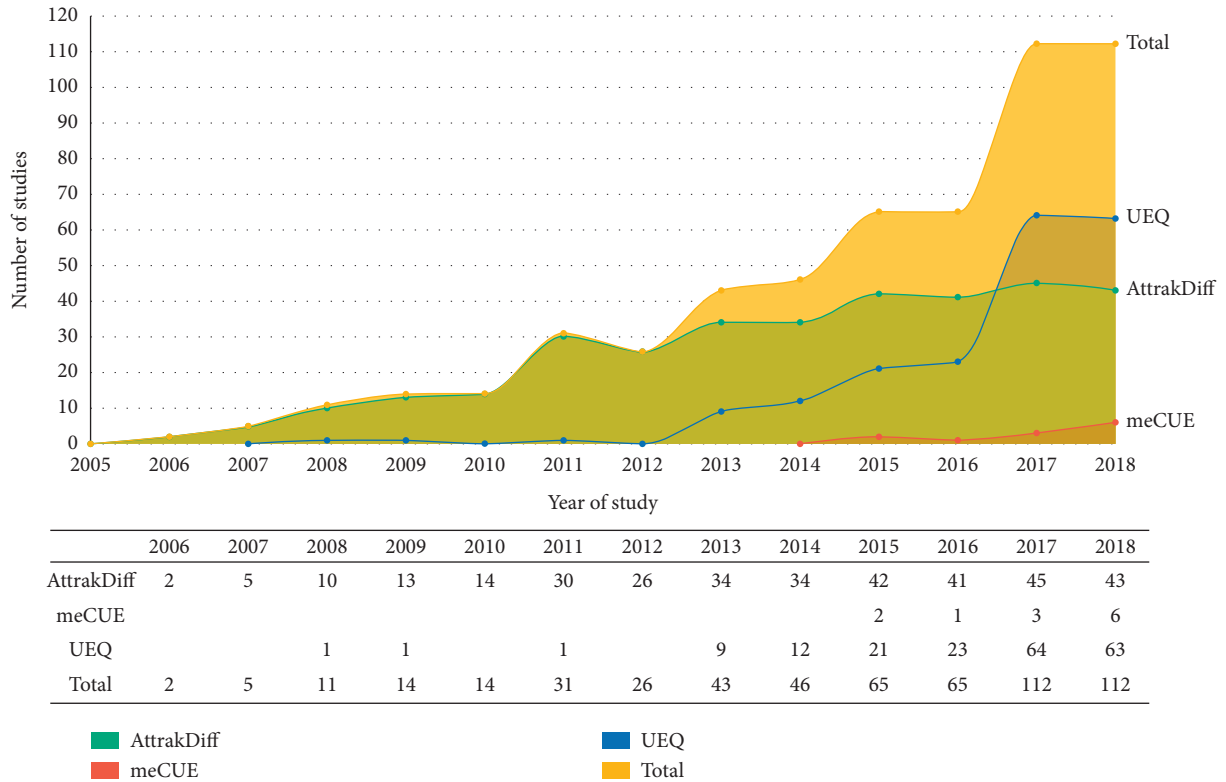


FIGURE 2: Total uses of standardized questionnaires by year, per questionnaire, and summarizing all three questionnaires.

TABLE 4: Total number of uses of standardized questionnaires by geographical region.

Questionnaire	Europe	Asia	North America	South America	Oceania	More than one region
AttrakDiff	303	7	10	8	5	7
UEQ	150	25	10	7	4	3
meCUE	10	1			1	
Total	463	33	20	15	10	10
Percentage	84.0%	6.0%	3.6%	2.7%	1.8%	1.8%

region with the most studies, with 463 studies out of 551 analyzed (84%). Europe is followed by Asia with 33 studies (6.0%), North America (20 studies, 3.6%), South America (15, 2.7%), and Oceania (10, 1.8%). Additionally, 10 studies (1.8%) were carried out in more than one region simultaneously, and 2 studies did not indicate the geographic location where the study was conducted.

Reviewing in detail the use of the questionnaires in Europe, the distribution of the three questionnaires corresponds to the global distribution of these. However, it is worth noting that in Asia, the UEQ questionnaire is significantly more used than the other two, with 76% of the uses (25 studies of 33 reviewed), while AttrakDiff only represents 21% (7 studies).

The large number of studies carried out in Europe would reflect the development and importance given to user experience in that region, particularly in Germany, which would have led to the creation of the three questionnaires taking place precisely in that country.

As shown in Figure 3, 247 studies of the 463 reported in Europe were conducted in Germany, representing a significant 53.3% of the studies in that continent. In countries

neighbouring Germany, such as Switzerland, Austria, the Netherlands, and France, there is also an important use of standardized questionnaires. Finland, although a little further away from Germany than those mentioned, contributes with 37 studies, representing 8.0% of the total. It is important to mention that the 100% indicated in Figure 3 corresponds to the 463 studies carried out in Europe and should not be confused with the percentage of total studies carried out worldwide (553).

Concerning other regions, Table 5 presents that Indonesia provides the largest number of studies carried out in Asia, with 11 of the 33 studies, corresponding to 33.3%, and using UEQ exclusively in all cases conducted in that country. In the same region, China adds 4 studies, Japan, Malaysia, and Taiwan add 3 studies each, the Philippines and South Korea, 2 studies each, and five more countries complete the list with 1 study. In South America, Brazil reports 8 of the 15 studies for 53.3% of that region, followed by Colombia with 3, Chile with 2, and Argentina and Peru with one study each. In Oceania, Australia contributes with 9 of the 10 studies (90.0%), while New Zealand contributes the tenth study.

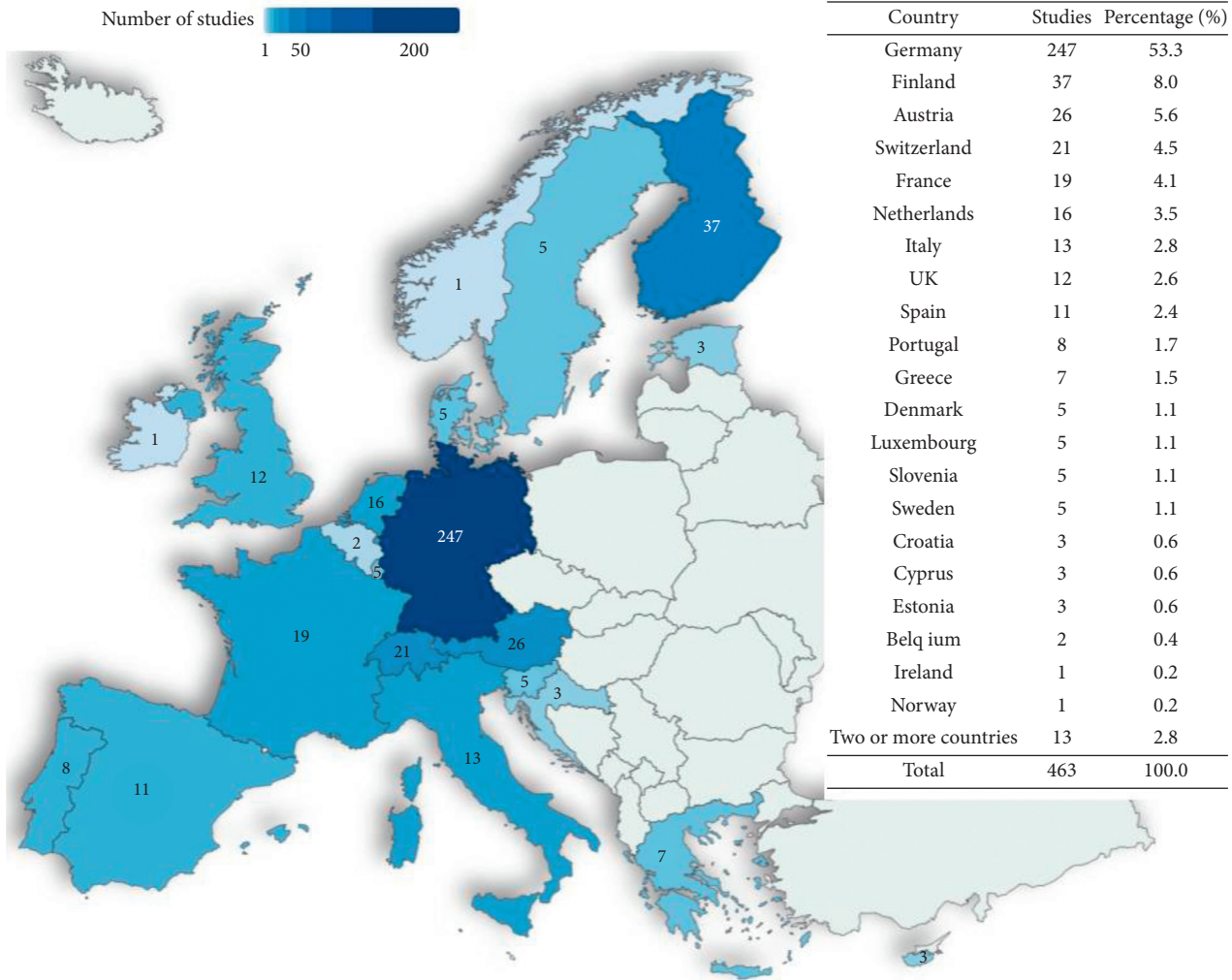


FIGURE 3: Total uses of standardized questionnaires in Europe.

In North America, the United States represents 50.0%, with 10 studies out of 20 reported in the region, followed by Mexico with 6 studies and Canada with 4. It is interesting to mention that of the 10 studies conducted in the United States, 7 were recently carried out, 3 in 2017, and 4 in 2018, which would seem to indicate that the interest on standardized questionnaires of researchers in that country is recent and that it could increase in the coming years.

**4.2. Number of Participants per Study (RQ2).** The number of participants in each of the reported studies ranged from 2 to 691 participants. It was not possible to identify the sample size in five studies, so the analysis presented is based on 548 studies. Figure 4 shows the median and quartile values for the number of participants in each study, for each standardized questionnaire, and for the aggregate value of the three questionnaires.

Regarding the aggregated data, the median is 20 participants per study, with the first quartile at 13 and the third quartile at 36. These values are similar for the three questionnaires if they are considered separately, with the medians of AttrakDiff and UEQ

in 21 participants and the median of meCUE in 19.5 participants. Quartiles one and three are also like those for the aggregated data, particularly for AttrakDiff and UEQ which are the questionnaires with the greatest number of studies. The fact that the general median for the three questionnaires is 20 participants and the median for AttrakDiff and UEQ is 21 could be influenced by the fact that the official AttrakDiff site has an online version where it is possible to collect information for up to 20 participants for free.

Analyzing whether the median number of participants per study is stable over the years, Figure 5 shows the values of the number of participants per study per year, where it can be seen that a median of around 21 participants has been maintained in recent years, with a slight growth in 2018 of 23 participants per study.

If the data corresponding to the journals studies are classified separately from the conference studies, 96 studies are identified as journal studies (17.5% of the total) and 452 as conference studies (82.5% of the total).

As shown in Figure 6, the median of participants per study in the journals studies is higher than the median of the



TABLE 5: Uses of standardized questionnaires by country in Asia, North America, South America, and Oceania.

	Country	Studies
Asia	Indonesia	11
	China	4
	Japan	3
	Malaysia	3
	Taiwan	3
	Philippines	2
	South Korea	2
	India	1
	Iran	1
	Singapore	1
	Sri Lanka	1
	Thailand	1
	USA	10
North America	Mexico	6
	Canada	4
	Brazil	8
South America	Colombia	3
	Chile	2
	Argentina	1
	Peru	1
Oceania	Australia	9
	New Zealand	1

studies of conferences (30 and 20 participants per study, respectively). This could mean that there is a filter in which studies published in journals should be supported by a larger number of participants per study. Also, significant is the fact that the third quartile of the journal studies is set at 70.5 participants, more than double the third quartile of the conference studies, which is set at 32. The conference studies, representing such a large proportion regarding the total number of studies, coincide almost exactly with the aggregated data of the journal and conference studies.

If the data are grouped by questionnaires, Figure 7 shows that AttrakDiff presents significantly higher values for journal studies, with a median of 40 participants compared to a median of 20 participants per study for conference studies.

Regarding UEQ studies, although the medians are similar between journal and conference studies, the range between the median and the third quartile of journal studies practically doubles the value of conference studies, indicating that the values are more scattered in the first case.

When comparing the data between questionnaires, the biggest difference is clear in the journal studies, with UEQ maintaining a median of 20 participants per study, while AttrakDiff showing a rather larger number, with a median of 42 participants per study. The meCUE questionnaire contributes with only 2 journal studies. It can be noted then that journal AttrakDiff studies would be mainly responsible for the upper median in journal studies that is shown in Figure 6. This behavior is not presented in the conference studies, where the values of the median and the first and third quartiles are remarkably similar between UEQ and AttrakDiff. Even the median of the ten meCUE conference studies matches that of the other two questionnaires.

In relation to the trend of participants per study for journal studies, Figure 8 shows that the median has been irregular over the years and based on a small number of studies. For the years 2017 and 2018, where there are more than 20 studies, the median is 20 and 26 participants per study, respectively, which could indicate an increasing trend, although these values are still between the median values for all studies combined (20 participants per study) and the median for journal studies, which is 30 participants per study.

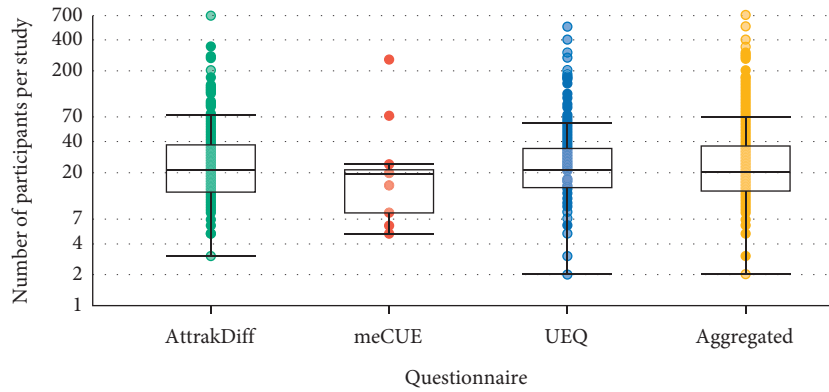
*4.3. Association with Other Evaluation Mechanisms (RQ3).* In 340 of the 553 studies reported (61.5%), in addition to applying the standardized UX evaluation questionnaire, researchers applied another evaluation instrument, while in the remaining 213 studies (38.5%), the standardized questionnaire was used as the only evaluation instrument. Of the 340 studies that complemented the evaluation with another instrument, 219 studies (64.4%) used one additional instrument, 88 studies (25.9%) used two, 28 studies (8.2%) used three, 4 studies (1.2%) used four, and 1 study (0.3%) used a total of five additional instruments.

Figure 9 shows the relationship between the three standardized questionnaires and other evaluation instruments used in the 553 studies. It is important to mention that since a study can include from zero to five additional instruments, the 340 studies that used at least one additional instrument represent 500 additional instruments. The 213 studies that did not use additional evaluations are considered as having one instrument each, which added to the 500 complementary instruments used, and gives a total of 713 instruments distributed in the graph.

In addition to showing the 213 studies that only used the standardized UX questionnaire as an evaluation instrument, Figure 9 presents the additional instruments most frequently used as a complement to the standardized UX questionnaires.

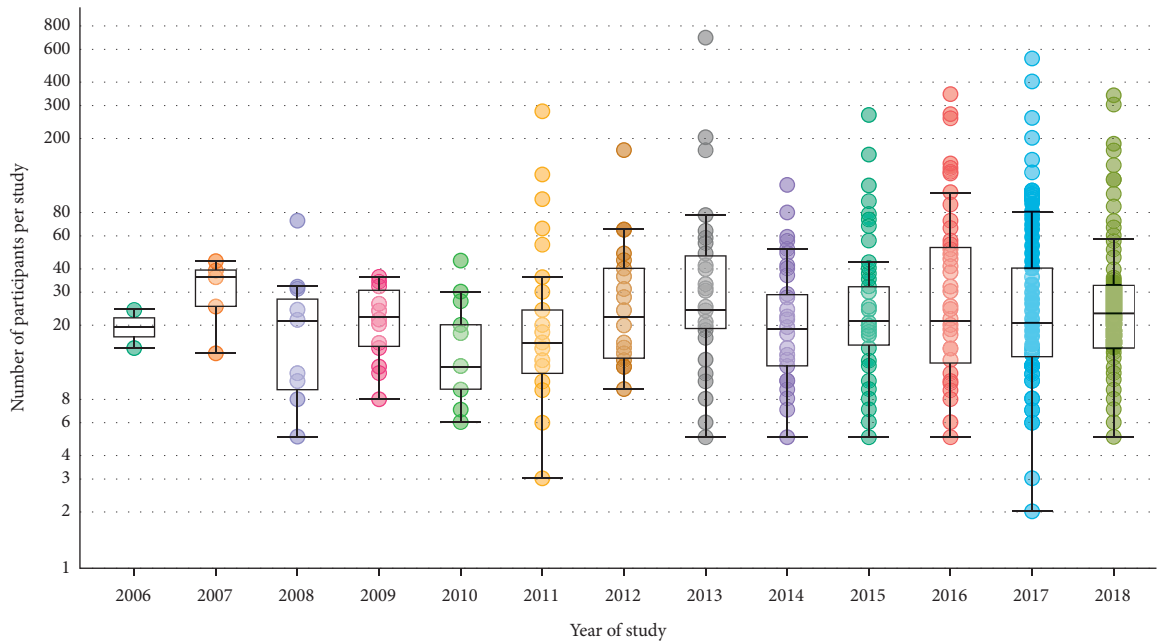
It is important to highlight that 120 studies applied the SUS (System Usability Scale) questionnaire, which demonstrates its strong positioning as a usability evaluation questionnaire. Other instruments used are self-designed questionnaires (72 studies), semistructured interviews (60 studies), NASA-TLX questionnaire (53 studies), PANAS (12 studies), the think aloud technique (11 studies), and 172 other instruments that were used in fewer than 10 studies each. Of these 172 instruments, 79 were used only once.

*4.4. Ubiquitous Computing and Ambient Intelligence (RQ4).* As part of the review of the 553 studies whose UX evaluation was performed with standardized questionnaires, 132 studies were identified that belong to trending topics on UbiComp and ambient intelligence. Initially, it is worth noting that the proportion of use of standardized questionnaires in these 132 studies varies in relation to the proportion of all studies. As given in Table 6, in studies with topics related to UbiComp and ambient intelligence, the



	AttrakDiff	meCUE	UEQ	Aggregated
Third quartile	37	21	34.25	36
Median	21	19.5	21	20
First quartile	12.75	8	14	13
Sample size	340	12	196	548

FIGURE 4: Number of participants per study by standardized questionnaires.

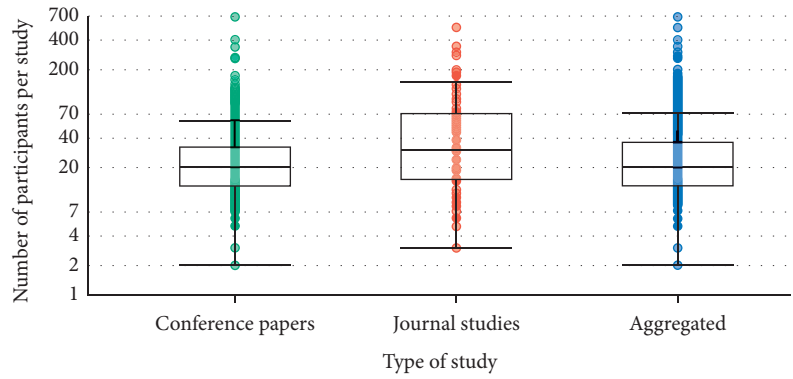


	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Third quartile	21, 75	39	27, 5	30, 5	20	24	40	46, 5	29	32	51, 75	40, 25	32, 5
Median	19, 5	36	21	22	12	16	22	24	19	21	21	20, 5	23
First quartile	17, 25	25	9	15, 25	9	11	13, 25	19	12	15, 5	12, 5	13, 5	15
Sample size	2	5	11	14	13	31	26	43	45	63	64	112	112

FIGURE 5: Number of participants per study per year.

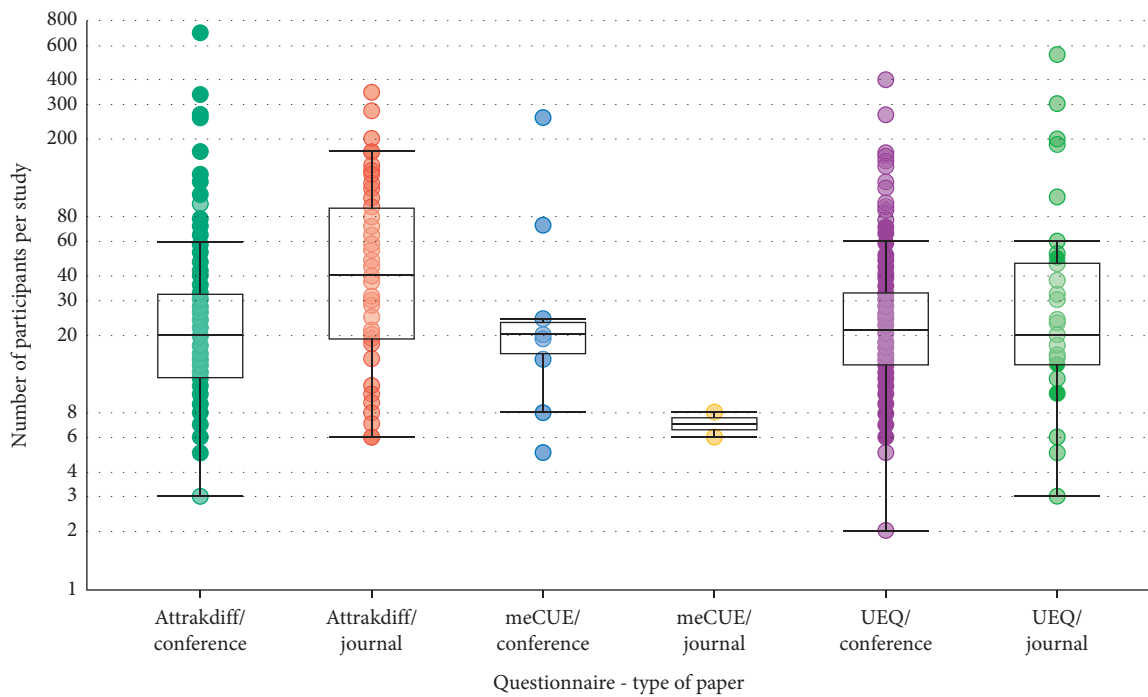
UEQ and meCUE questionnaires increase their proportion of use with respect to the proportion present in all the studies (from 36.2% to 40.2% for UEQ and from 2.2% to 3.8% for meCUE). Obviously, the use of the AttrakDiff questionnaire decreases to compensate for this fact.

Seven topics related to UbiComp and ambient intelligence were identified. These topics are shown in Figure 10, as well as the number of studies in each category. These topics can be associated with environments, so they would correspond to ambient intelligence categories,



	Conference studies	Journal studies	Aggregated
Third quartile	32	70.5	36
Median	20	30	20
First quartile	13	15	13
Sample size	452	96	548

FIGURE 6: Number of participants per study by the type of the study.



	Attrakdiff/ conference	Attrakdiff/ journal	meCUE/ conference	meCUE/ journal	UEQ/ conference	UEQ/ journal
Third quartile	32	88	23	7.5	32.5	46
Median	20	40	20	7	21	20
First quartile	12	19	16	6.5	14	14
Sample size	283	57	10	2	159	37

FIGURE 7: Number of participants per study by standardized questionnaires.

except for the topic called IoT and wearable sensors, which is defined around those specific technologies. This topic was included to highlight the important use of these

technologies, and it should be noted that a study can be classified in more than one topic. For this reason, being in the IoT and wearable sensors category does not inhibit a

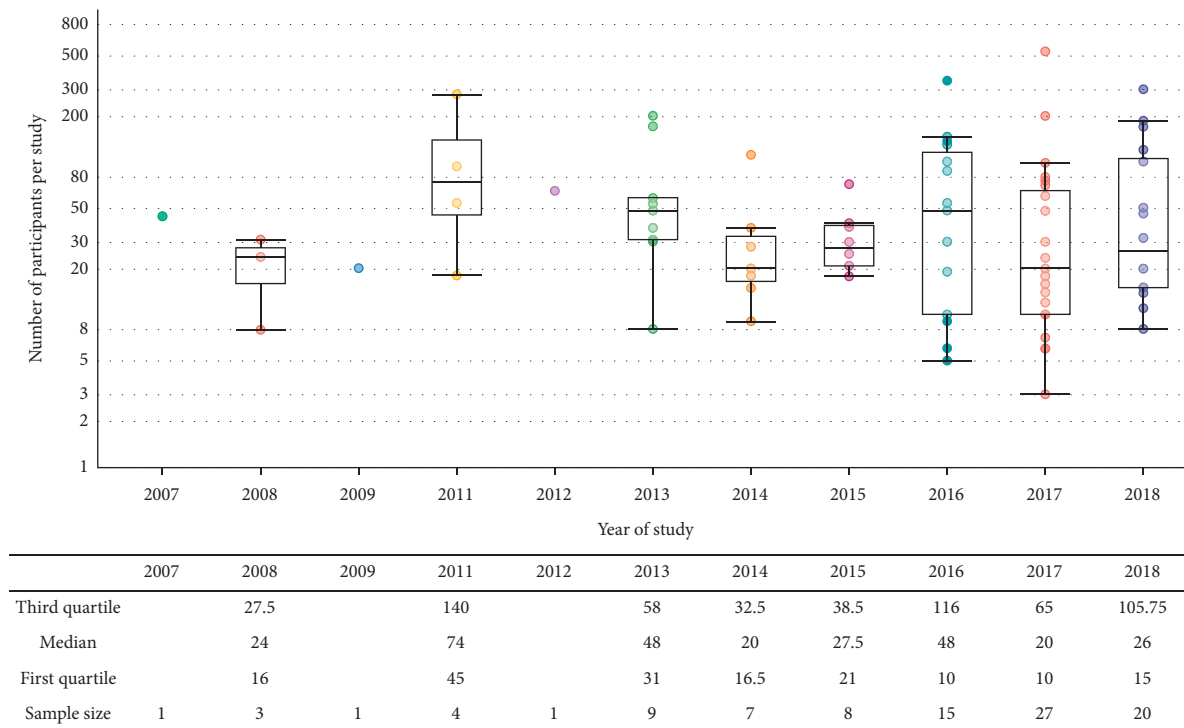


FIGURE 8: Number of participants per study in journal studies by year.

study from appearing in another category, such as smart environments for health.

The topic most studies identified is IoT and wearable sensors with 38 studies. Some of these studies are also related to other topics as shown in Figure 10. For example, the use of IoT and wearable to monitor people activity, especially for the elderly, to assess people's posture or the rehabilitation of the way of walking after neurological events (also part of the topic smart environment for health), to provide spatial information about the proximity of objects to people walking (indoor positioning and navigation), or through gloves or tactile wrist watches to motorcyclists (in-vehicle information systems), among others. Our analysis also identified studies in which IoT and wearables are used to provide information in other contexts such as work or leisure, by means of bracelets that provide light information, rings, or smart watches. It is interesting to note that most studies on this topic are implemented using nontraditional interfaces. This point will be detailed in Section 4.5.

The second topic with the most appearances is in-vehicle information systems, with 36 studies. In this category, we include a set of studies mainly from the automotive industry, where interactions are made between the car and the driver or between the motorcycle and the rider, through the projection of information on the windshield or helmet visor or through haptic interfaces on the steering wheel or gloves and touchless gesture interfaces for car driving, among others.

The third category corresponds to 27 studies related to smart cities, smart homes, and other human-ambient interaction. The studies on smart cities correspond to the

efficient use of energy, the use of renewable energies, and the report of urban incidents or people engagement, among others. In the case of smart homes, studies were identified proposing interfaces within homes or seeking to identify emotions such as loneliness or anxiety in people inside the household. Studies that use geospatial data to provide information about urban environments and other human-ambient interactions such as displays in museums, in public art installations, or in buildings that are part of the architectural heritage and that vary according to different elements such as the proximity of people, were also included in this category.

After that, 22 studies related to intelligent transportation systems were identified, dealing with different topics, among which it is worth highlighting shipment tracking, route planning, social interaction between vehicles, public transportation navigation information, and urban pedestrian navigation. This category also overlaps with the in-vehicle information systems category in 8 of the studies, since the interaction inside the vehicle has to do with the environment outside the vehicle, especially with regard to safety. For example, providing the pilot with environmental information related to safety in order to improve vehicle driving and, consequently, improve overall traffic.

The next category in number of studies is smart environments for health, with 18 studies that include voice interfaces for home care and home communication services, support for elderly and assistance in independent living, and biofeedback scenarios as visualization of heart activity. In this category were also included studies on medical tools such as robot assistant surgery, X-ray imaging, needle guidance, and poststroke rehabilitation using virtual reality,



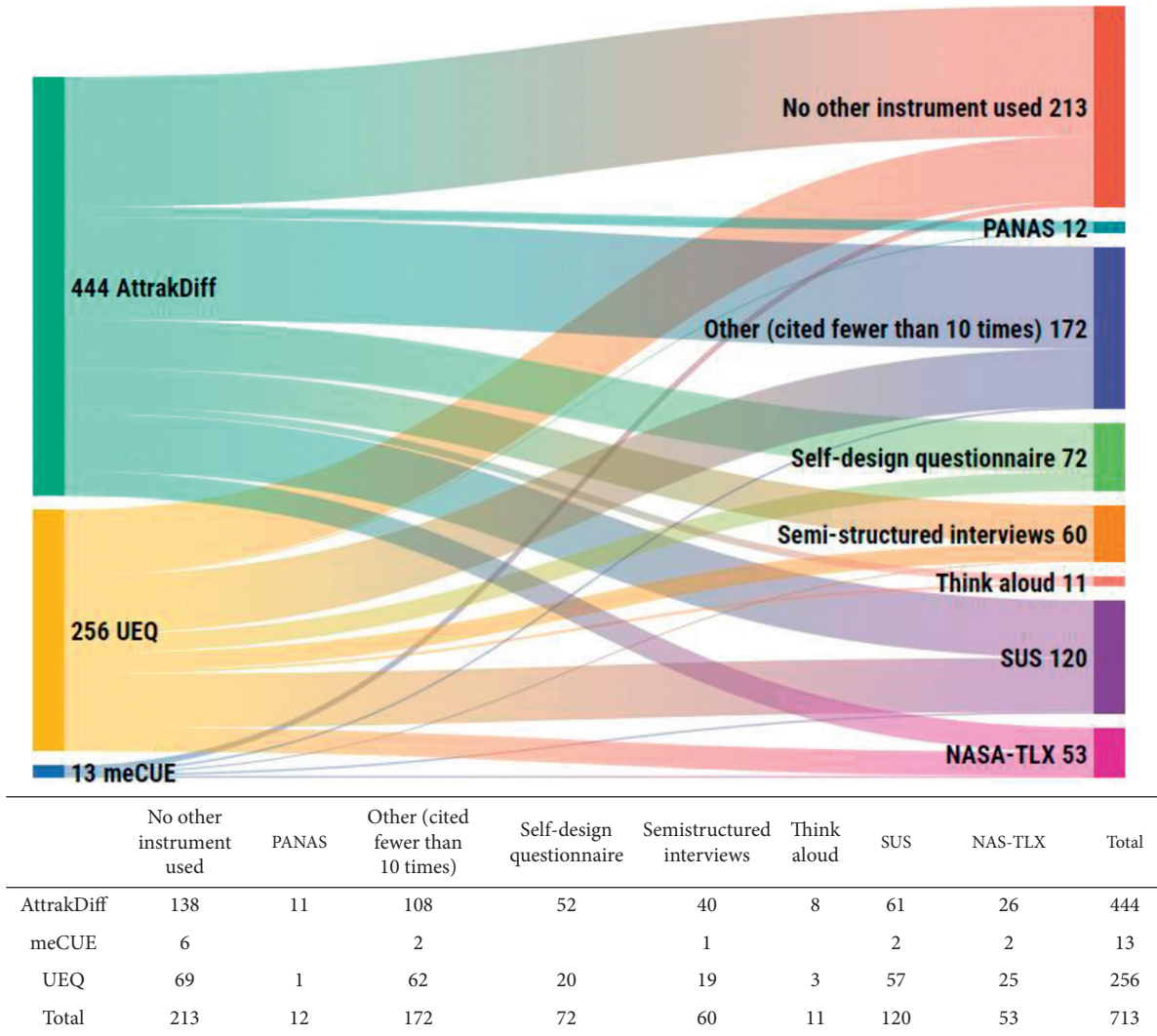


FIGURE 9: Additional instruments used in combination with the three standardized questionnaires.

TABLE 6: Percentage of standardized questionnaires for all studies and for studies with topics related to ubiquitous computing and ambient intelligence.

	All studies	Percentage	UbiComp and ambient intelligence studies	Percentage
AttrakDiff	341	61.7	74	56.1
meCUE	12	2.2	5	3.8
UEQ	200	36.2	53	40.2
Total	553	100.0	132	100.0

among others. This topic shares studies with smart homes, such as those mentioned above, on identifying emotions such as loneliness or anxiety within the household.

There are two categories with the fewest studies identified. First, a category that groups studies related to indoor positioning navigation together with proposals for navigation on virtual environments (9 studies in total), and finally, the category Internet of people that includes 5 studies of interactions between people in contexts such as public transport, public art installations, or activities such as locating services in emergencies or sharing music in urban settings.

The use of the AttrakDiff, UEQ, and meCUE in these seven topics is shown in Figure 11, where the UEQ questionnaire surpasses the uses of AttrakDiff for studies on IoT and wearable sensors, even though AttrakDiff has higher proportions in UbiComp and ambient intelligence studies in general, as given in Table 6. An element that could be contributing to this fact would be the year in which the study was carried out.

As shown in Figure 12, IoT and wearable sensors studies are grouped in recent years, with a median of 2017, in which UEQ has taken the lead to the AttrakDiff questionnaire, so it could be a general trend and not a specific trend from studies

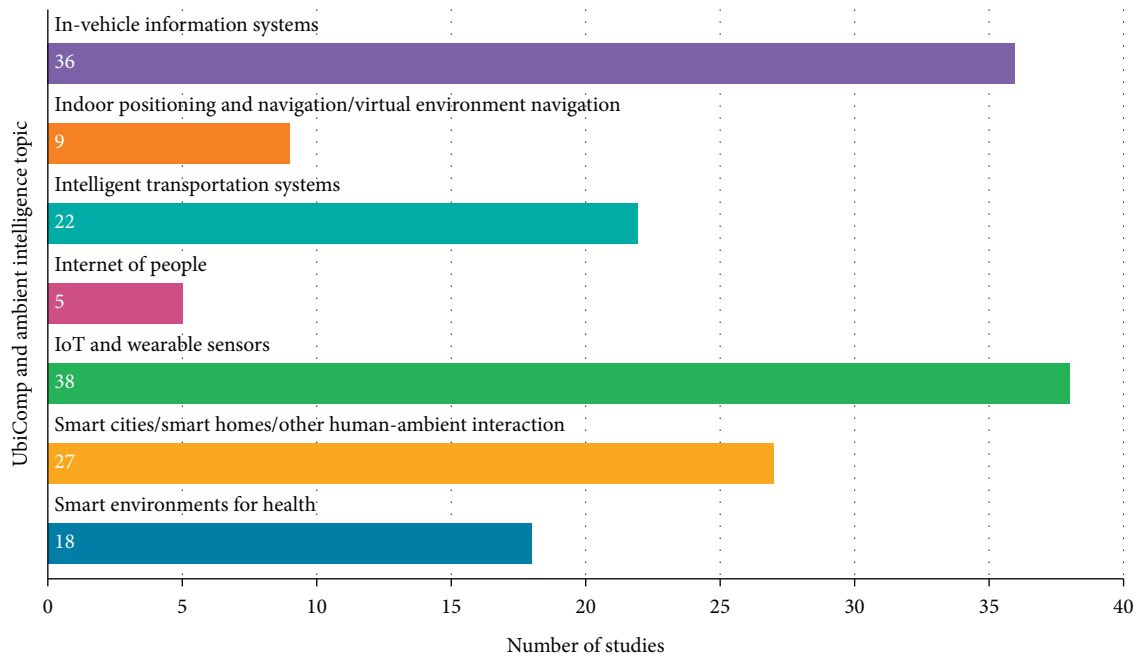


FIGURE 10: Number of ubiquitous computing and ambient intelligence studies evaluated with standardized questionnaires classified by topic.

on IoT and wearable sensors. Another interesting case is the topic smart environments for health, where the same number of studies is presented for AttrakDiff and UEQ, and an important participation of meCUE also appears, with 22% of the studies on this topic. As in the previous topic, the year of the study could be a factor that influences this fact.

The opposite case occurs in the topics in-vehicle information systems, smart cities/smart homes/other human-ambient interaction, and Internet of people, where the use of AttrakDiff is greater than UEQ, even above the general proportions (66.7%, 70.4%, and 80.0%, respectively). These data are presented in Table 7.

In the case of in-vehicle information systems, this tendency to use AttrakDiff predominantly does not seem to be related to the year of the study, since, as shown in Figure 12, there are numerous recent studies on this topic, and the median of this topic is only slightly lower than the median of all studies. The majority use of AttrakDiff in studies on smart cities/smart homes/other human-ambient interaction and Internet of people could indeed be influenced by a trend of these studies in years prior to 2015, where AttrakDiff was the predominant questionnaire.

The studies on the topic indoor positioning and navigation/virtual environment navigation use more AttrakDiff than UEQ with a slightly lower proportion than the general proportion (55.6%), which could be due to the influence of the year of the study or the fact that in this topic, only 9 studies were identified. Finally, the intelligent transportation systems studies use more AttrakDiff than UEQ as a UX evaluation instrument, but in a similar proportion to the data shown for all studies.

**4.5. Nontraditional Interfaces (RQ5).** A set of 181 studies, out of the total of 553 studies included in the systematic literature review, uses nontraditional interfaces. As in the previous section referring to studies related to UbiComp and ambient intelligence, studies with nontraditional interfaces show a greater use of UEQ and meCUE questionnaires with respect to the proportion present in all 553 studies and with an even greater difference (from 36.2% to 43.1% for UEQ and from 2.2% to 4.4% for meCUE). As Table 8 presents, the use of the AttrakDiff questionnaire again decreases to compensate for this fact from 61.7% to 52.5% in this case.

To classify these 181 studies, 12 categories of nontraditional interfaces were defined, following the guidelines presented by Kortum [35] and Karray et al. [34], although not strictly. Figure 13 shows the categories identified and the number of studies found in each category.

As it can be seen, studies with virtual reality interfaces are the most numerous, represented by 30 studies. Studies with gesture interfaces is next with 28 studies, among which no studies with gestures on smart phones or tablets were included, as they were considered traditional interfaces. The nontraditional visual interfaces category, 20 studios, consists of interfaces that present information by visual means, such as head-up displays or projections on helmet visors, as well as light information on nontraditional interfaces such as armbands. Haptic interfaces are presented in 19 of the studies analyzed. Then, 16 studies related to eye tracking or gaze detection were identified. The next group in number of uses is tangible interfaces with 14 studies. Next, 13 studies are identified where the interface is carried out with robots. Multimodal interfaces appear in 12 studies. In this group, in which the task is

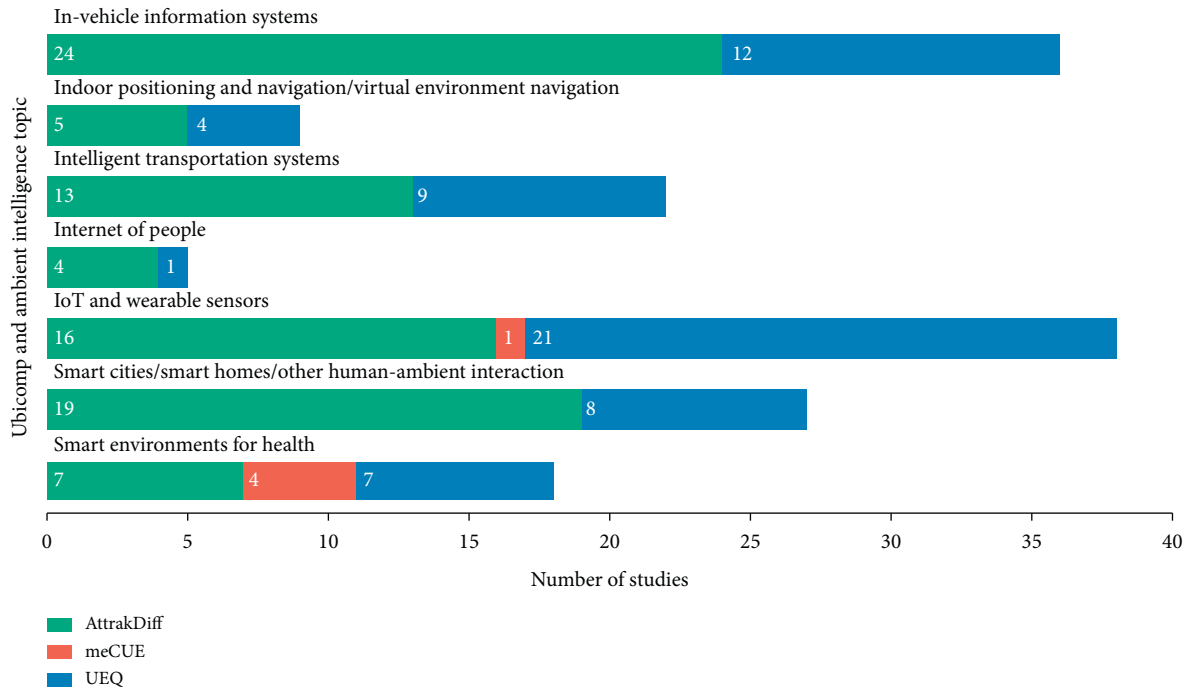


FIGURE 11: Number of uses of standardized questionnaires in UbiComp and ambient intelligence studies by topic.

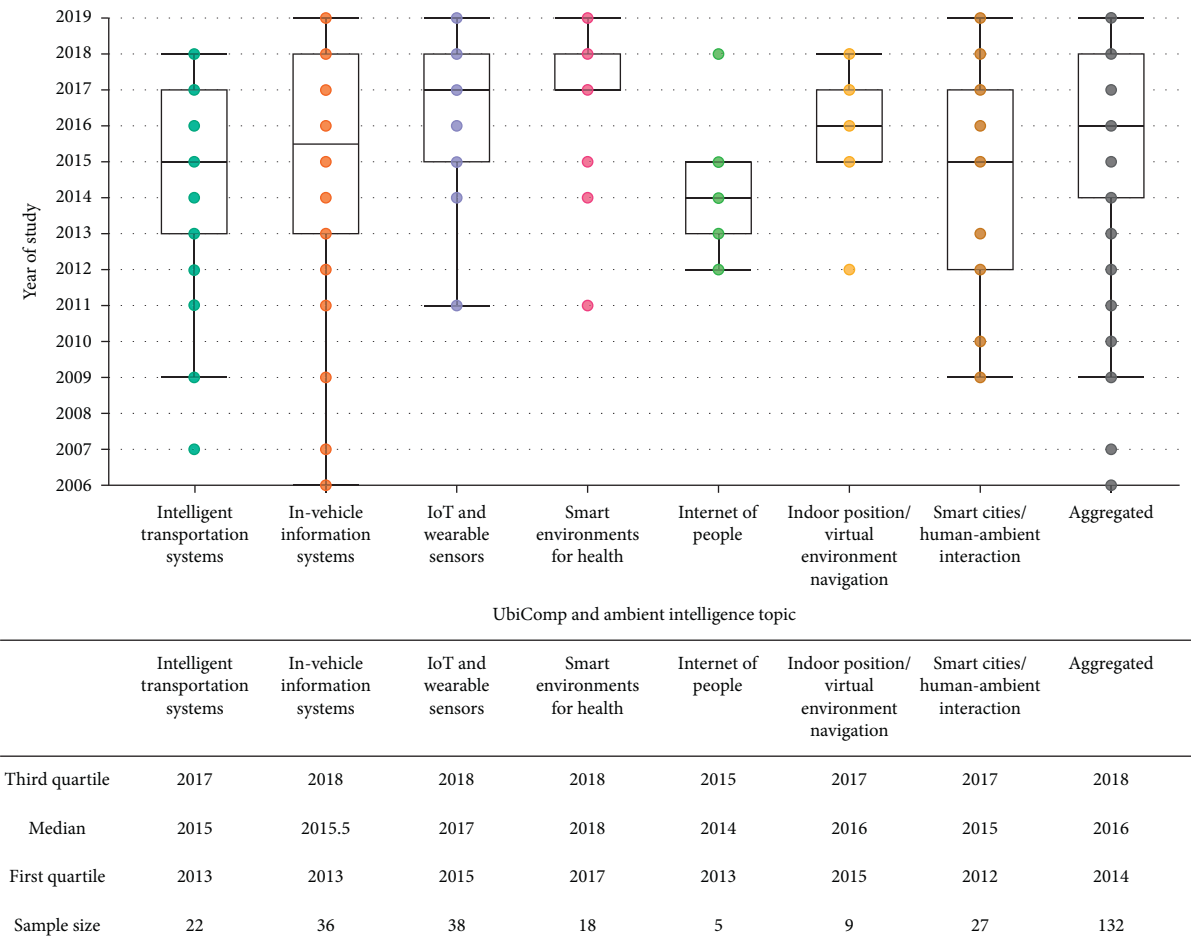


FIGURE 12: Ubiquitous computing and ambient intelligence studies evaluated with standardized questionnaires by year.

TABLE 7: Percentage of standardized questionnaires for UbiComp and ambient intelligence studies by topic.

	AttrakDiff	%	meCUE	%	UEQ	%	Total
Intelligent transportation systems	13	59.1	—	—	9	40.9	22
In-vehicle information system	24	66.7	—	—	12	33.3	36
IoT and wearable sensors	16	42.1	1	2.6	21	55.3	38
Smart environments for health	7	38.9	4	22.2	7	38.9	18
Internet of people	4	80.0	—	—	1	20.0	5
Indoor positioning/virtual environment navigation	5	55.6	—	—	4	44.4	9
Smart cities/smart homes/other human-ambient interaction	19	70.4	—	—	8	29.6	27

TABLE 8: Percentage of standardized questionnaires in studies with nontraditional interfaces.

	All studies	Percentage (%)	Nontraditional interfaces studies	Percentage (%)
AttrakDiff	341	61.7	95	52.5
meCUE	12	2.2	8	4.4
UEQ	200	36.2	78	43.1
Total	553	100.0	181	100.0

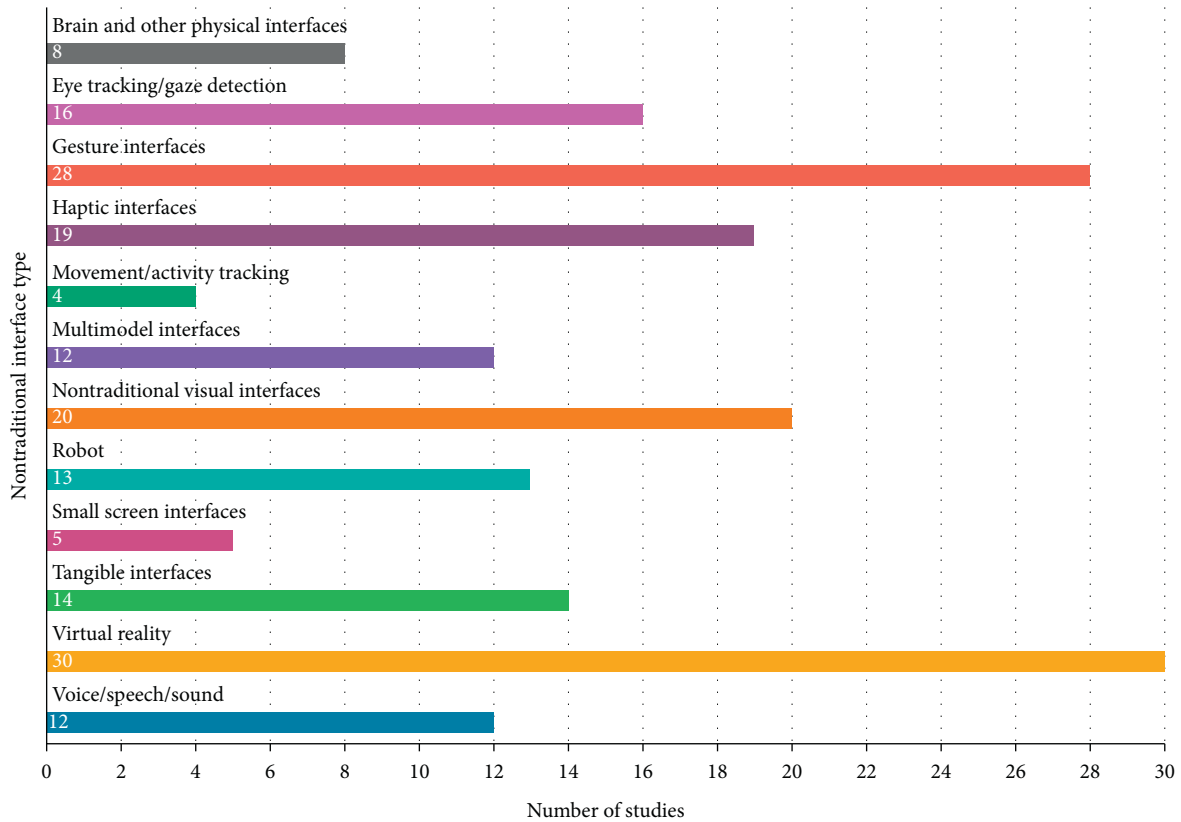


FIGURE 13: Number of studies with nontraditional interfaces by interface type.

performed by more than one interface, studies related to virtual reality, which is also multimodal, were not included, since they are such a large group that it justified a category in itself. Furthermore, multimodal category includes both mutually inclusive and mutually exclusive multimodal interfaces, according to Kortum classification [35], given that the vast majority corresponds to mutually inclusive interfaces, where the interaction with the system is carried out through several interfaces simultaneously.

Studies with interfaces related to voice, speech, and sound were grouped into one category with 12 studies. This grouping covers the auditory, speech, and interactive voice response categories presented in [35]. Eight studies presented systems with brain interfaces and other physiological measures not included in the haptic category. Finally, 5 studies present smart screen interfaces, mostly smart watches, and 4 studies deal with systems with movement or activity trackers, implemented with bands or wrist devices.



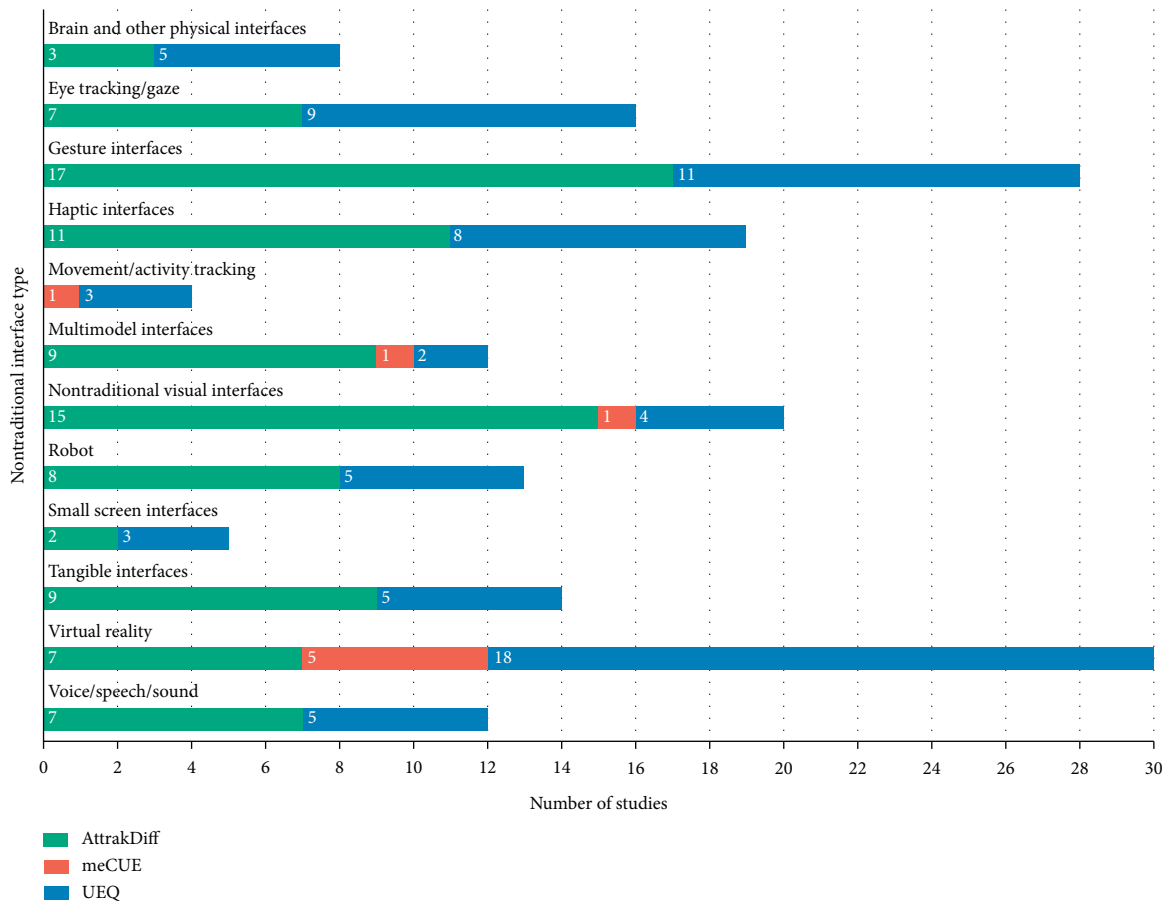


FIGURE 14: Number of uses of standardized questionnaires in studies with nontraditional interfaces by the type of interface.

Based on the categories defined above, Figure 14 shows that the UEQ questionnaire surpasses the uses of AttrakDiff for studies with various types of interfaces, including virtual reality, eye tracking/gaze detection, movement/activity tracking, small screen interfaces, and brain and other physical interfaces.

An element that could be contributing to this relationship would be the year the study was conducted. This situation is shown in Figure 15, where studies with virtual reality, movement/activity tracking, and small screen interfaces accumulate most of their studies in recent years, where UEQ has taken the lead over the other two questionnaires.

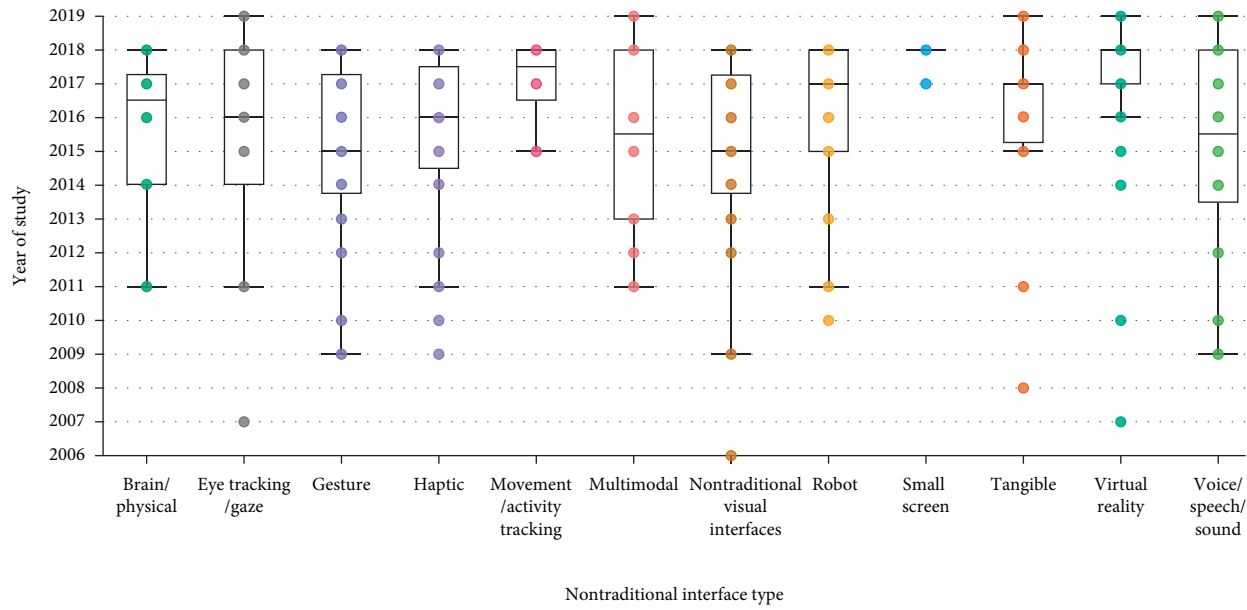
This relationship with the year is not so evident for studies with eye tracking/gaze detection and brain and other physical interfaces, which could indicate a preference of researchers to use UEQ in this type of study, especially in eye tracking/gaze detection, where the number of studies is greater and could be more significant.

On the other hand, studies with tangible, multimodal, and nontraditional visual interfaces show a marked preference to use AttrakDiff, with numbers well above the relationship presented in Table 8. In these three categories, the predominant use of AttrakDiff could be determined by the year of the study, given that in all three cases, the median of the studies is just above 2015, as given in Table 9.

Finally, studies with haptic, gesture, voice/speech/sound, and robot interfaces use more AttrakDiff than UEQ as a UX evaluation instrument, but in a proportion similar to all studies, it appears that these types of interfaces do not influence which questionnaire to use.

Another point to analyze is the use of nontraditional interfaces in UbiComp and ambient intelligence studies that used standardized questionnaires as an instrument for UX evaluation. Table 10 presents this relationship. In total, 86 studies of the UbiComp and ambient intelligence topics used nontraditional interfaces. It is worth mentioning that, as indicated in the previous sections, there are studies that belong to more than one UbiComp and ambient intelligence topic, but the categorization of nontraditional interfaces is only one category per study.

As expected, the topic IoT and wearable interfaces is the one with the most uses of nontraditional interfaces, with 34 studies identified. Of these 34, haptic interface is the most frequent, along with gesture and smart screens present in smart watches. The topic in-vehicle information system used 24 nontraditional interfaces, being the second largest group. These interfaces include 13 nontraditional visual interfaces, which are mainly made up of systems that present information on different surfaces of the vehicle such as in windshields or in motorcycle helmets. This relationship between the topic in-vehicle information



	Brain/physical	Eye tracking/gaze	Gesture	Haptic	Movement/activity tracking	Multimodal	Nontraditional visual interfaces	Robot	Small screen	Tangible	Virtual reality	Voice/speech/sound
Third quartile	2017.25	2018	2017.25	2017.5	2018	2018	2017.25	2018	2018	2017	2018	2018
Median	2016.5	2016	2015	2016	2017.5	2015.5	2015	2017	2018	2017	2018	2015.5
First quartile	2014	2014	2013.75	2014.5	2016.5	2013	2013.75	2015	2018	2015.25	2017	2013.5
Sample size	8	16	28	19	4	12	20	13	5	14	30	12

FIGURE 15: Studies with nontraditional interfaces by year.

TABLE 9: Percentage of standardized questionnaires in studies with nontraditional interfaces.

	AttrakDiff	%	meCUE	%	UEQ	%	Total
Tangible interfaces	9	64.3	—	—	5	35.7	14
Nontraditional visual interfaces	15	75.0	1	5.0	4	20.0	20
Haptic interfaces	11	57.9	—	—	8	42.1	19
Movement/activity tracking	—	—	1	25.0	3	75.0	4
Gesture interfaces	17	60.7	—	—	11	39.3	28
Eye tracking/gaze detection	7	43.8	—	—	9	56.3	16
Voice/speech/sound	7	58.3	—	—	5	41.7	12
Small screen interfaces	2	40.0	—	—	3	60.0	5
Virtual reality	7	23.3	5	16.7	18	60.0	30
Multimodal interfaces	9	75.0	1	8.3	2	16.7	12
Brain and other physical interfaces	3	37.5	—	—	5	62.5	8
Robot	8	61.5	—	—	5	38.5	13

TABLE 10: Relationship between UbiComp and ambient intelligence studies with nontraditional interfaces for studies evaluated with standardized questionnaires.

	In-vehicle information systems	Indoor positioning navigation	Intelligent transportation systems	Internet of people	IoT and wearable sensors	Smart cities/Ambient interaction	Smart environment for health
Brain/physical	—	—	—	—	3	1	2
Eye tracking/gaze detection	3	—	2	—	—	—	—
Gesture	4	2	1	1	5	3	—

TABLE 10: Continued.

	In-vehicle information systems	Indoor positioning navigation	Intelligent transportation systems	Internet of people	IoT and wearable sensors	Smart cities/ Ambient interaction	Smart environment for health
Haptic	—	1	1	—	7	—	—
Movement/ activity tracking	—	—	—	—	3	—	1
Multimodal	2	—	2	—	3	2	—
Nontraditional visual interfaces	13	1	4	—	4	1	1
Robot	—	—	—	—	1	—	3
Small screen	—	—	1	—	5	—	—
Tangible	1	—	1	—	2	1	—
Virtual reality	—	2	—	—	1	—	7
Voice/speech/ sound	1	1	—	—	—	2	2

systems with nontraditional visual interfaces is consistent in the use of the AttrakDiff questionnaire, which we previously discussed seems to be well positioned in the automotive industry.

Finally, it is important to highlight the significant use of virtual reality interfaces in smart environments for health, with 7 uses out of 16 studies on this topic, in which virtual reality is being used, for example, in medical devices for organ visualization or diagnostic support.

## 5. Limitations

The results of the presented study may have been affected by the selection process carried out by the group of researchers, which could be influenced by their human characteristics. Having used a large group of researchers poses a challenge to the consistency of the inclusion criteria and characterization of the studies. Cross validations were performed to reduce biases.

Another point to mention is that only four digital databases were used for collecting the studies. Although the number of studies analyzed is significant, future studies could consider including other sources.

## 6. Conclusions

This study presents the results of the systematic literature review conducted to classify and compare the uses of the standardized questionnaires AttrakDiff, UEQ, and meCUE in academic studies. This review was conducted around five research questions (Section 3.1) that encompass the purposes pursued in this SLR. These questions were answered extensively in Section 4. Some of the more interesting results found for those research questions are presented.

Results show that the use of standardized questionnaires has increased year after year, starting in 2006, where first studies were published describing their use. Throughout these years, the most used questionnaire is AttrakDiff, which coincides with the fact of being the first questionnaire to be created. However, since 2017, the UEQ questionnaire has far surpassed AttrakDiff in number of uses.

As for the geographical context, the standardized questionnaires have been used more extensively in Europe than in the rest of the world, followed by Asia. And within Europe, Germany greatly exceeds the rest of the European countries. It should be noted that the three questionnaires, although their original version is in German, were quickly translated into English, so that their use could be more widespread. Despite this and the United States being one of the technological leaders of the world, few studies using standardized questionnaires are reported in that country. It should be noted, however, that the 10 studies reported in the United States correspond to the years 2017 and 2018, which could indicate that the use of standardized questionnaires will increase in the coming years.

Regarding the number of participants of the studies, this study shows that the median for the aggregated data of the three questionnaires is 20 participants per study, while the values for the first and third quartiles are 13 and 36 participants, respectively. This information is similar for the two most used questionnaires: AttrakDiff and UEQ if their data are analyzed individually. However, if the journal studies are analyzed separately (17.5% of the total studies) from the conference studies (82.5% of the total), it can be seen that the median of the journal studied rises to 30 participants per study, which is well above the 20 participants of the general median and the median for conference studies. This could mean that journal studies require evaluations with larger sets of participants to be accepted or that the same authors collect data from more participants when evaluating studies to be published in journals assuming that this would increase the possibilities of publication. It is interesting to note that this increase in the number of participants is mainly being provided by studies with the AttrakDiff questionnaire, which is greater but only in journal studies. In conference studies, both AttrakDiff and UEQ studies present medians of 20 and 21 participants, respectively, which correspond to the overall median.

It is also worth mentioning that 38.5% of the studies reviewed used the standardized UX evaluation questionnaire as the only evaluation instrument, which would indicate that the investigators have confidence in the instrument, limited time or resources to design and apply the evaluation, or a

combination of these factors. The remaining 61.5% of primary studies (340 studies) used between one and five complementary instruments, among which the SUS usability questionnaire stands out, reported in 120 of the studies analyzed.

Regarding the use of standardized questionnaires to evaluate UX in UbiComp and ambient intelligence studies, there is an important set of studies identified: 132 out of the total 553 studies, about 24% of the total. Among the topics with the most studies categorized are IoT and wearable interfaces, in-vehicle information systems, smart cities/smart homes/other human-ambient interaction, and intelligent transportation systems. In this set of 132 studies, the proportion of uses of the UEQ questionnaire increases in relation to the proportion of the 553 studies (from 36.2% to 40.2%), and the proportion of meCUE also increases (from 2.2% to 3.8%). Some topics favour UEQ or AttrakDiff over general proportions, which could be influenced by the date of the study, given that since 2017, studies using UEQ have surpassed those using AttrakDiff. In some cases, however, it could be due to a preference for one or the other, such as the sustained use of AttrakDiff in the automotive industry on the topic of in-vehicle information systems. On the other hand, the combination of the increasing number of UEQ uses in trend topics such as IoT and wearable sensors could lead to the consolidation of UEQ as a standard questionnaire for evaluating UX in studies on this topic.

About studies that include solutions or systems with nontraditional interfaces, 181 studies were identified, which represent about 33% of the 553 studies that used standardized questionnaires as a UX evaluation instrument. These 181 studies were classified into 12 different types of interfaces, being virtual reality and gesture interfaces the most numerous. Similar to studies classified as UbiComp and ambient intelligence topics, studies with nontraditional interfaces show an increase in the uses of UEQ and meCUE compared to the total of studies and even in a higher proportion than that found by topic: UEQ goes from 36.2% to 43.1% and meCUE goes from 2.2% to 4.4%, doubling its participation. Again, the date of the studies may represent an important factor in the increased use of UEQ and meCUE. This fact, added to the significant difference in the uses of UEQ for studies with virtual reality and eye tracking/gaze detection, could make the use of UEQ the preferred standardized questionnaire for UX evaluation of this type of interface. In the case of virtual reality, the incursion of meCUE as an evaluation instrument is noteworthy, with 5 studies out of the 30 identified for this interface, so it should be followed carefully in the coming years to determine if it manages to dispute the prevalent position UEQ has in studies with virtual reality. In other nontraditional interface cases where the difference in favour of UEQ is also significant, such as movement/activity tracking, small screen, and brain and other physical interfaces, the number of studies is still small to determine if there is a marked preference.

On the other hand, studies with tangible interfaces, nontraditional visual interfaces, and multimodal interfaces (other than virtual reality) show a strong preference of researchers to use AttrakDiff. In all three cases, the date of

the study may be a factor to consider. However, studies catalogued as nontraditional visual interfaces include an important set of studies with displays in windshields and motorcycle helmets, related to the topic of the in-vehicle information system, of which we previously commented that AttrakDiff seems well positioned in the automotive industry.

Finally, combining UbiComp and ambient intelligence topic and nontraditional interface, it is noted that studies related to IoT and wearable sensors were the ones that used the most nontraditional interfaces, among studies that used standardized questionnaires as UX evaluation instrument.

## Data Availability

The data used to support the findings of this study are included within the article, and the complete bibliography of the 548 studies analyzed in the full-text review is available at [http://citic.ucr.ac.cr/sites/default/files/SLR\\_References\\_IDO\\_2020.pdf](http://citic.ucr.ac.cr/sites/default/files/SLR_References_IDO_2020.pdf).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was funded by ECCI and CITIC at the University of Costa Rica (834-B4-412 and 326-B9-107).

## Supplementary Materials

The complete bibliography of the 548 papers analyzed in the full-text review is available at [http://citic.ucr.ac.cr/sites/default/files/SLR\\_References\\_IDO\\_2020.pdf](http://citic.ucr.ac.cr/sites/default/files/SLR_References_IDO_2020.pdf). (*Supplementary Materials*)

## References

- [1] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (UEQ) in different evaluation scenarios," in *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer International Publishing, Crete, Greece, January 2014.
- [2] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries-NordiCHI '10*, p. 521, Reykjavik, Iceland, October 2010.
- [3] D. Wallach, J. Conrad, and T. Steimle, "The UX metrics table: a missing artifact," in *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer International Publishing, Vancouver, BC, Canada, May 2017.
- [4] 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.



- [5] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, "User experience white paper: bringing clarity to the concept of user experience," in *Proceedings of the Result from Dagstuhl Seminar on Demarcating User Experience*, Helsinki, Finland, September 2011.
- [6] E. Law, V. Roto, A. Vermeeren, J. Kort, and M. Hassenzahl, "Towards a shared definition of user experience," in *Proceedings of the CHI'08 Extended Abstracts on Human Factors in Computing Systems (CHI EA'08)*, Association for Computing Machinery, New York, NY, USA, April 2008.
- [7] ISO, *ISO DIS 9241-210:2010 Ergonomics of Human-System Interaction*, International Standardization Organization (ISO), Geneva, Switzerland, 2010.
- [8] C. Rohrer, "When to use which user-experience research methods," 2014, <https://www.nngroup.com/articles/which-ux-research-methods/>.
- [9] C. Lallemand and G. Gronier, *Méthodes de design UX. 30 méthodes fondamentales pour concevoir des expériences optimales*, Éditions Eyrolles, Paris, France, 2nd edition, 2018.
- [10] G. Gronier, C. Lallemand, and A. Chauvet, *Mesurer la formation de la première impression d'une interface à l'aide du test des 5 secondes; Huitième Colloque de Psychologie Ergonomique (EPIQUE)*, The Luxembourg Institute of Science and Technology (LIST), Aix-en-Provence, France, 2015.
- [11] C. Lallemand and V. Koenig, "How could an intranet be like a friend to me?" in *Proceedings of the European Conference on Cognitive Ergonomics 2017 - ECCE 2017*, Umeå, Sweden, September 2017.
- [12] J. Baumgartner, A. Sonderegger, and J. Sauer, "No need to read: developing a pictorial single-item scale for measuring perceived usability," *International Journal of Human-Computer Studies*, vol. 122, pp. 78–89, 2019.
- [13] Y. Forster, S. Hergeth, F. Naujoks, and J. F. Krems, "How usability can save the day - methodological considerations for making automated driving a success story," in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 278–290, Toronto, ON, Canada, September 2018.
- [14] J. Klammer and F. W. G. van den Anker, *Design, User Experience, and Usability: Users, Contexts and Case Studies*, Springer International Publishing, Berlin, Heidelberg, Germany, 2018.
- [15] C. L. B. Maia and E. S. Furtado, *A Systematic Review about User Experience Evaluation. In Design, User Experience, and Usability: Design Thinking and Methods*, Springer International Publishing, Berlin, Heidelberg, Germany, 2016.
- [16] A. C. Ten and F. Paz, "A systematic review of user experience evaluation methods in information driven websites," in *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 492–506, Springer International Publishing, Vancouver, BC, Canada, February 2017.
- [17] R. A. Likert, *Technique for the Measurement of Attitude*, Archives of Psychology, New York, NY, USA, 1932.
- [18] E. C. Osgood, "The Nature and measurement of meaning," *Psychological Bulletin*, vol. 49, pp. 197–237, 1952.
- [19] M. Hassenzahl, M. Burmester, and F. A.D. Koller, "Ein Fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität," in *Proceedings of the Mensch & Computer 2003: Interaktion in Bewegung*, pp. 187–196, Stuttgart, Germany, April 2003.
- [20] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," *Computer Science*, vol. 5298, pp. 63–76, 2008.
- [21] M. Minge and L. Riedel, "meCUE – ein modularer fragebogen zur erfassung des nutzungserlebens [meCUE - a modular questionnaire for capturing the user experience]," *Mensch und Computer*, vol. 9, pp. 89–98, 2013.
- [22] M. Hassenzahl, "The effect of perceived hedonic quality on product appealingness," *International Journal of Human-Computer Interaction*, vol. 13, pp. 481–499, 2001.
- [23] M. Hassenzahl, "The thing and I: understanding the relationship between user and product," in *Funology : From Usability to Enjoyment*, M. A. Blyth, Ed., pp. 1–12, Kluwer Academic Publishers, New York, NY, USA, 2003.
- [24] M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in human-technologyinteraction," *International Journal of Psychology*, vol. 42, no. 4, pp. 253–264, 2007.
- [25] M. Seel, *Ubiquitous Computing. In Encyclopedia of the Sciences of Learning*, Springer, Boston, MA, USA, 2012.
- [26] M. Weiser, "The computer for the 21 st century," *Scientific American*, vol. 265, no. 3, pp. 94–105, 1991.
- [27] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J. Burgelman, *Scenarios for Ambient Intelligence in 2010. IST Advisory Group Final Report*, European Commission, Brussels, Belgium, 2001.
- [28] P. Remagnino, H. Hagra, N. Monekosso, and S. Velastin, "Ambient Intelligence. a gentle introduction," in *Ambient Intelligence*, P. Remagnino, G. L. Foresti, and T. Ellis, Eds., Springer, New York, NY, USA, 2005.
- [29] W. Weber, J. Rabaey, and E. Aarts, *Ambient Intelligence*, Springer, Berlin, Heidelberg, Germany, 2005.
- [30] J. C. Augusto, "Ambient intelligence: basic concepts and applications," in *Software and Data Technologies. ICISOFT 2006. Communications in Computer and Information Science*, J. Filipe, B. Shishkov, and M. Helfert, Eds., Springer, Berlin, Heidelberg, Germany, 2008.
- [31] M. Gams, I. Yu-Hua Gu, A. Härmä, A. Muñoz, and V. Tam, "Artificial intelligence and ambient intelligence," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 1, pp. 71–86, 2019.
- [32] M. Friedewald, O. Da Costa, Y. Punie, P. Alahuhta, and S. Heinonen, "Perspectives of ambient intelligence in the home environment," *Telematics and Informatics*, vol. 22, no. 3, pp. 221–238, 2005.
- [33] M. Bick and T. Kummer, "Ambient intelligence and ubiquitous computing," in *Handbook on Information Technologies for Education and Training. International Handbooks on Information Systems*, H. H. Adelsberger, P. J. M. Kinshuk, and D. G. Sampson, Eds., Springer, Berlin, Heidelberg, Germany, 2008.
- [34] F. Karray, M. Alemzadeh, J. Abou Saleh, and M. Nours Arab, "Human-computer interaction: overview on state of the art," *International Journal on Smart Sensing and Intelligent Systems*, vol. 1, pp. 137–159, 2008.
- [35] P. Kortum, *HCI beyond GUI: Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces*, Elsevier/Morgan Kaufmann Publishers Inc., California, CA, USA, 1st edition, 2008.
- [36] A. Liberati, D. G. Altman, J. Tetzlaff et al., "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions," *Explanation and Elaboration*, vol. 62, pp. e1–e34, 2009.

- [37] M. Minge, M. Thüring, I. Wagner, and C. V. Kuhr, *The meCUE Questionnaire: A Modular Tool for Measuring User Experience*, pp. 115–128, Springer, Cham, Switzerland, 2017.
- [38] M. Minge and M. Thüring, “The MeCUE questionnaire (2.0): meeting five basic requirements for lean and standardized UX assessment,” in *Proceedings of the HCI International 2018*, pp. 451–469, Las Vegas, NV, USA, July 2018.
- [39] I. Wagner and M. Minge, “The gods play dice together: the influence of social elements of gamification on seniors’ user experience,” in *Proceedings of the Communications in Computer and Information Science*, Springer International Publishing, Los Angeles, CA, USA, May 2015.
- [40] K. Pollmann, V. Sinram, and N. Fronemann, “Design. Can we distinguish pragmatic from hedonic user experience qualities with implicit measures?” in *Proceedings of the HCI International 2018*, pp. 509–527, Las Vegas, NV, USA, July 2018.
- [41] M. Möttus, D. Lamas, and L. Kuk, “Aesthetic categories of interaction: aesthetic perceptions on smartphone and computer,” in *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems*, pp. 1249–1256, Prague, Czech Republic, September 2017.
- [42] M. A. Ali, “Data-Driven knowledge acquisition system: an end-to-end knowledge engineering process for generating production rules,” *IEEE Access*, vol. 6, pp. 15587–15607, 2018.
- [43] S. Gebhardt, “An evaluation of a smart-phone-based menu system for immersive virtual environments,” in *Proceedings of the 2014 IEEE Symp. 3D User Interfaces*, pp. 31–34, Minneapolis, MN, USA, March 2014.
- [44] S. Gebhardt, S. Pick, F. Leithold, B. Hentschel, and T. Kuhlen, “Extended pie menus for immersive virtual environments,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 644–651, 2013.