Exploring Usability and User Experience Evaluation Methods: A Tertiary Study

Geremias Corrêa[1][a], Roberto Pereira[2][b], Milene Selbach Silveira[3][c] and Isabela Gasparini[1][d]

[1] *Universidade do Estado de Santa Catarina (UDESC), Joinville, Brazil*

[2] *Universidade Federal do Paraná (UFPR), Curitiba, Brazil*

[3] *Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil*

Keywords: Tertiary Study, Evaluation Methods, Usability, User Experience.

Abstract: Usability and User Experience (UX) evaluation methods have important roles in business and scientific spheres, effectively pinpointing areas for enhancement across a broad spectrum of applications. Primary and secondary scientific studies investigating these methods are relevant and provide a panorama of different domains. While providing macro views on the topic is necessary, tertiary studies are still uncommon. This paper fills this gap by presenting a tertiary study conducted through a systematic search methodology, following Petersen's guidelines. Studies indexed by Scopus, IEEE Xplore, and ACM search engines were considered, resulting in 487 retrieved studies, from which 36 were deemed relevant, and another 7 studies were added through a snowballing search strategy. From the selected studies, methods, domains of application, and considerations for the inclusion of accessibility in studies, among other information, were identified and discussed. Results revealed Questionnaires as the prevalent method in these studies, Brazil and Indonesia as the leading countries in authorship of publications, and Observation, Inspection, and Inquiry as the most common category for methods. These results suggest a prevalence of well-structured methods, generally with lower costs and application times, revealing space for further investigation.

1 INTRODUCTION

Usability and User Experience (UX) evaluation methods are commonly applied to anticipate and reveal problems that may affect the quality of user interaction and interface. Its application can occur during and after the development of a given product. The evaluation methods vary in format, structure, goal, target audience, user profiles, and application domain.

Understanding the applicability of each method and choosing the most suitable is not easy as one must consider their application cost, time, profile of the target audience, effectiveness in a given context, and viability, among other issues. Furthermore, changes in technology, new application domains, and characteristics of the target audience are factors that require evaluation methods to be updated and revisited.

We first looked for secondary studies on the topic to identify the panorama of the literature on Usability and UX evaluation methods. Focusing on systematic mappings and reviews covering one or more methods and their application in primary studies, 27 reviews, ranging between 2012 and 2021, were found – the threshold date at the time of the initial search. On the one hand, in an exploratory analysis, we identified a significant recurrence of secondary studies in recent years, showing a saturation of secondary studies.

On the other hand, we found no tertiary study cataloging and analyzing these secondary studies, offering a macro and structured overview of the current knowledge in the field. A tertiary study enables us to identify, understand, and organize relevant information about these studies, such as what methods are applied, what is evaluated, which study domains are addressed, which countries have investigated the subject, as well as understanding the evaluation methods used, their categorizations, forms of application and other relevant characteristics.

A systematic mapping of the literature was carried out to prepare for this tertiary work, substantiating and defining the research questions and the

[a] https://orcid.org/0009-0007-8948-3297
[b] https://orcid.org/0000-0003-3406-3985
[c] https://orcid.org/0000-0003-2159-551X
[d] https://orcid.org/0000-0002-8094-9261

Corrêa, G., Pereira, R., Silveira, M. and Gasparini, I.
Exploring Usability and User Experience Evaluation Methods: A Tertiary Study.
DOI: 10.5220/0012606100003690
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 2*, pages 357-368
ISBN: 978-989-758-692-7; ISSN: 2184-4992
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

search method, followed by the definition and application of the inclusion and exclusion criteria. The elaborations emerged according to defined guidelines(Petersen et al., 2008; Petersen et al., 2015).

From the initial search on Scopus, IEEE Xplore, and ACM Digital Library, 487 studies were retrieved, and 36 were included after the inclusion and exclusion criteria were applied. Another 7 studies were selected by using the snowballing technique. From the selected studies, we drew information that reveals relevant findings about the topic, such as the most investigated and applied (type of) methods, the most addressed domain context, how in-depth these secondary analyses have been, whether accessibility has been an agenda in the evaluation of these works, and other relevant research questions. The paper is organized as follows. Section 2 introduces the fundamental concepts of this research. Section 3 presents and discusses related work. Section 4 details the systematic mapping process carried out, the research questions, the search process, and inclusion and exclusion criteria. Section 5 explores the results obtained and answers the research questions, while Section 6 discusses the results obtained. and summarizes the main findings of the research, also bringing perspectives for future work.

2 CONCEPTS

In the rapidly evolving world of technology and digital design, usability and user experience (UX) have emerged as elements for the success of any product or service (Soares et al., 2022). On the one hand, Usability refers to the ease with which a user can navigate and interact with a product or system, aiming for efficiency, effectiveness, and satisfaction in a specific context of use. On the other hand, UX takes a broader perspective, encompassing the entire spectrum of a user's interaction with a product, including emotional, psychological, and behavioral responses.

Designers employ various evaluation methods to ensure these elements meet user needs and expectations. These methods range from user testing, where real users interact with the product in controlled environments, to heuristic evaluations, where experts use established guidelines to assess usability. Surveys and analytics also play a role, providing quantitative data on user satisfaction and behavior. Together, these concepts and methods form the backbone of creating Human-centric digital products that are not only functional but also provide an engaging user experience.

2.1 Usability and User Experience (UX)

Usability refers to the ease with which users can interact with a product or system to achieve their goals effectively and efficiently while having a satisfactory experience.

Usability encompasses different factors, being also defined by (Barbosa et al., 2021; Nielsen, 1994):

1. Ease to learn: the system needs to be simple to learn so the user can quickly start interacting;
2. Efficient to use: the system needs to be efficient in use so that once learned, the user has a high level of productivity;
3. Ease to remember: the system needs to be uncomplicated to remember so that the user, when using it again after a certain time, does not have to learn it again;
4. Few errors: an error is defined as an action that does not lead to the expected result and that should be minimized. There may be contexts of simple errors, which only delay the user, as well as catastrophic errors, which have impacts blocking the user in their action.
5. Satisfaction: users must like the system, that is, it must be pleasant so that the user is content when using it.

The term user experience is used to describe a lot of meanings, including the usability of hedonic resources, the measurement of affect, or the user experience in interactions (Nagalingam and Ibrahim, 2015). User experience includes all user emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors, and achievements that occur before, during, and after use(ABNT, 2011).

UX includes cognitive, sociocultural, and affective aspects - positive aspects of users' experience in their product interaction, such as aesthetic experience or desire to reuse the product. It covers all aspects of the user experience with the system, involving all aspects of end users' interaction with your company, services, and products (Norman, 2014). Rogers et al., 2013, consider that while usability is concerned with the criteria of efficiency, effectiveness, and satisfaction, the user experience addresses the quality of the experience. Thus, the concepts differ in the ways and means to achieve an objective. The application of these aspects and those defined for usability can be measured using evaluation methods.

2.2 Evaluation Methods

The evaluation of software is an important activity during the entire development and post-development process of a product (Buse et al., 2011). Evaluative methods emerge to measure aspects with different approaches. HCI evaluations are necessary for validating the interface according to user requirements, verifying difficulties in its use, identifying interaction barriers, and comparing alternative interface designs (da Silva Osorio et al., 2008). There are several methods for evaluating interfaces, which have different characteristics depending on the context they address. It is necessary to understand these characteristics to identify which methods are suitable for application according to the study's target goals.

HCI evaluation methods are usually categorized according to the form of evaluation: one way to categorize them is by classifying them as Inquiry, Inspection, and Observation (Barbosa et al., 2021). Inquiry methods allow the author's interpretation and analysis based on the responses of those evaluated, *e.g.*, use of questionnaires, interviews, and focus groups. Inspection methods allow evaluation by experts to predict future user experiences, *e.g.*, heuristic evaluations, and cognitive walkthroughs. Finally, Observation methods are usually characterized by data recording, allowing real problems to be identified during the evaluator's experience using the system, *e.g.*, eye tracking, and usability tests.

3 RELATED WORK

Secondary studies have revealed different aspects of HCI evaluation methods, mainly analyzing how evaluation methods were applied and the form and depth of applications. However, a tertiary study was not found to organize and provide a macro-view of the literature. Therefore, some secondary studies directly related to this study are discussed next.

In Fernandez et al., 2012, the authors selected 18 from 206 retrieved studies published between 1996 and 2009, analyzing the most applied usability evaluation methods in the websites. The authors identified a need for more research on empirical methods, including quality methods for analysis, and showed a need for better standardization in measuring the methods' effectiveness. The authors, however, did not investigate an in-depth analysis of the identified methods.

Prietch et al., 2022 investigated Usability and UX evaluation methods for automated sign language processors indexed by ACM DL, IEEE Xplore, Science Direct, SpringerLink, Scopus, Web of Science, Taylor and Francis Online, and Google Scholar. The authors selected 37 studies published from 2015 to 2020, categorizing them into generation, recognition, and translation – which are relevant terms for the investigated context. The Questionnaire method was the most applied method, followed by Prototyping, Experiments, and Usability Testing.

Yanez-Gomez et al., 2017 evaluated 187 studies published between 2003 and 2015 in the IEEE Xplore, ACM DL, and Web of Knowledge bases, focusing on usability methods applied to the Serious Games domain. The authors identified the Questionnaire as the most commonly used method. They also noted that Serious Games on health and learning require special attention from usability evaluation and offered an opportunity for further research.

Considering another related study, Maia and Furtado, 2016 analyzed the general application domain of UX evaluation methods. Analyzing 25 primary studies published between 2008 and 2016, available at IEEE Xplore, ACM DL, and Science Direct bases, they identified the Questionnaire method as the most applied one (84.00%) and that sensory measurements are rarely used, probably due to higher costs and complex application process.

The studies in this section do not cover both usability and UX evaluation methods, nor do they offer a comprehensive analysis of the literature across all application domains. However, they present pertinent findings and illustrate the diversity of research published in recent years, underscoring the feasibility of conducting a tertiary review.

4 SYSTEMATIC MAPPING OF LITERATURE

This study presents a systematic mapping of the literature on Usability and UX evaluation methods in the form of a tertiary study to identify and structure the methods covered by secondary studies. For Kitchenham and Charters, 2007, a tertiary study is necessary in a domain with a sufficient number of secondary studies, so evaluating them using a methodology similar to secondary studies becomes valid. Therefore, a systematic mapping was designed to conduct this work, which allows the categorization of a large portion of studies in the literature.

This work adopts Petersen's methodology (Petersen et al., 2008; Petersen et al., 2015). The mapping protocol and its application were built by the first author of this work. All the authors analyzed the results, discussed the findings, and participated in the analysis and writing. The mapping process is described in the following subsections.

4.1 Research Questions

The Research Questions (RQ) of our tertiary review are presented as a Main Research Question (MRQ) and Secondary Research Questions (SRQ) as follows:
MRQ: What Usability and User Experience evaluation methods have been used in the literature?

- **SRQ1:** Does history analysis reveal the prominence of specific methods? If so, which ones?
- **SRQ2:** How many primary studies were analyzed?
- **SRQ3:** Is there a classification standard for the evaluation methods used?
- **SRQ4:** In which application domains and subdomains are these evaluation methods inserted?
- **SRQ5:** Is accessibility a factor considered in secondary studies? If so, in what way?

In MRQ, the analysis of which methods were found is defined as a substantial point of this study. With this, the aim was to obtain secondary works on this issue that seek the clear application of methods in primary works and measure such applications, synthesizing and reflecting on them. Because of this greater importance, other issues are defined as secondary. Despite this, SRQ1 is complementary to MRQ, seeking to analyze the possible dominance of a certain method found in secondary studies.

Regarding the other SRQ, we aimed to obtain relevant information for the analysis. The number of studies analyzed allows us to see the average number of selected studies and understand the scope and possible depth of the studies.When looking for methods classification strategies, we can clearly define and apply each method. When investigating the domains and subdomains of application of the methods, the most common contexts of application are observed, and whether any specific area or subarea can outline any behavior or expectation.

Finally, accessibility is examined due to its importance - integrating accessibility and human values is a cornerstone for creating equitable digital experiences. This approach goes beyond mere compliance with standards; it embodies a deeper understanding of human values such as empathy, respect, and dignity. By prioritizing these values, designers can create interfaces that not only meet the functional requirements of users but also resonate with them on a personal and emotional level. Therefore, we investigate accessibility as a transversal factor, observing whether the Usability and UX methods consider and discuss it on a broader level. However, we expect its presence in selected studies not to be expressive, partly because of our research's focus and because accessibility is usually evaluated with specific methods.

4.2 Search Process

After defining the research questions, we determined the search string for retrieving relevant studies. The process consisted of an exploratory search for different arguments that could meet the initial requirement: to return secondary works that addressed the mapping of Usability and User Experience evaluation methods in primary studies. The quality of each string argument was determined based on the relevance analysis of the first 10 studies found in each database application. After classifying and refining the search string, arranged into 4 search arguments, the following definition was adopted:

(”systematic review” OR ”systematic mapping” OR ”literature review”) AND (”user experience” OR usability) AND (techniques OR methods) AND (evaluation)

The first argument was related to the type of study, seeking to obtain secondary studies in a general way. The second argument represents the two possible contexts of studies: Usability and User Experience. In the third argument, the terms ”techniques” and ”methods” were considered due to combining both terms for the results obtained. We chose to conduct our search using plural terms to retrieve mappings and systematic reviews that explore various evaluation methods. In our fourth argument, we made the term 'evaluation' mandatory. This decision was based on our preliminary results, which showed a tendency to exclude studies that did not focus on measuring the evaluation of methods or measured only a single technique, diverging from our intended scope of research.

Once the search argument was developed, the search bases were selected based on the work of Buchinger et al., 2014, which presents an analysis of the performance of different research bases. Therefore, the search string was applied in the bases, and the number of results obtained and their efficiency were analyzed. The bases that presented the best results were IEEE Xplore, Scopus, and ACM DL.

Works were selected based on title, abstract, or keyword match. For Scopus, the initial search only filtered studies in the Computer Science area. Filtering was necessary as Scopus indexes studies from different sources and offers a filter by area. On the other hand, ACM DL and IEEE did not include any initial filtering, as they are already dedicated to Computer Science and related areas.

Following, we defined the inclusion and exclusion criteria to include only works that can answer the research questions (Petersen et al., 2008). The order of the criteria is related to the order in which they were applied. So, the analysis was first carried out by observing the inclusion criteria and, after that, the exclusion criteria, resulting in the Table 1.

Table 1: Inclusion and Exclusion criteria definition.

Inclusion criteria
IC1 - Publication year between 2012 and 2022
IC2 - Studies completed available by the university access
IC3 - Non-duplicate studies
IC4 - Studies longer than three pages
IC5 - English language studies
IC6 - Original publication studies
IC7 - Studies from journals or scientific events
Exclusion criteria
EC1 - Non-secondary studies
EC2 - Studies that do not measure UX or usability evaluation methods in primary studies

The covered period, from 2012 to 2022, in IC1, occurs due to the results found in the initial searches, also considering avoiding studies with results that are outdated or that would put too much stress on subsequent analyses. In IC2, we selected studies that the university portal can fully access. In IC3, non-duplicate studies about those already obtained through other databases were added. IC4 aimed to eliminate studies that could be too short and could not have enough content to be characterized as secondary studies. The target language in IC5 was considered only English to be as close to fair for all countries due to its internationalization. IC6 aims to obtain original publications, avoiding studies that could be reviews in other events and journals. Lastly, IC7 defines only journals or scientific events as the base source to avoid studies with low support and criticality.

Regarding the exclusion criteria, EC1 is necessary to define whether it is a secondary study. The study's scope also measures the application of Usability and UX methods defined by EC2.

With the analysis of the 487 studies initially retrieved, 36 studies were selected after the inclusion and exclusion criteria were applied, as Table 2 describes. Considering the 36 studies obtained, 13 come from Scopus, 13 from IEEE, and 10 from ACM DL. Each criterion had an impact on the exclusion of at least 1 study. The most determining criterion was EC2, eliminating 48.46% from the total number of excluded eliminated. This criterion eliminated the majority of works that, despite reaching the last stage, had little to do with the target scope. Despite this, some recognized works were eliminated for dealing with evaluation methods in a way that was not expected. Studies without Usability or User experience methods measurements of the application, such as in-depth studies or discussions around the topic, were disregarded.

Table 2: Analysis of the remaining bases and studies after applying each criterion.

Step	Scopus	IEEE	ACM	Total
Initial	154	103	230	487
IC1	141	84	194	419
IC2	124	84	194	402
IC3	110	79	186	375
IC4	110	79	176	365
IC5	108	78	170	356
IC6	108	77	170	355
IC7	108	77	168	353
EC1	91	62	119	272
EC2	13	13	10	36

After the selection, a backward snowballing process (Wohlin, 2014) was conducted to obtain more works that could be relevant for the research but that, for some reason, had not been reached by the search string. Therefore, another 7 papers were added. The summary of the selected studies is shown in the Table 3.

Table 3: List of 43 selected secondary studies.

ID	Reference	Base	Context	Primary Studies
S1	(Fernandez et al., 2012)	IEEE	Usability	18
S2	(Araujo et al., 2014)	IEEE	Usability	12
S3	(Paz and Pow-Sang, 2014)	IEEE	Usability	274
S4	(Zapata et al., 2015)	Scopus	Usability	22
S5	(Feather et al., 2016)	Scopus	UX	21
S6	(Paz and Pow-Sang, 2015)	IEEE	Usability	228
S7	(Yanez-Gomez et al., 2017)	Scopus	Usability	187
S8	(Ellsworth et al., 2017)	Scopus	Usability	120
S10	(Khodambashi and Nytrø, 2017)	Scopus	Usability	20
S11	(Yerlikaya and Onay Durdu, 2017)	Scopus	Usability	53
S12	(Zarour and Alharbi, 2017)	Scopus	UX	114
S13	(Ansaar et al., 2020)	IEEE	Usability	19
S14	(Saare et al., 2020)	Scopus	Usability	24
S15	(Weichbroth, 2020)	IEEE	Usability	75
S16	(Sheikh et al., 2021)	IEEE	Usability	15
S17	(Almazroi, 2021)	Scopus	Usability	62
S18	(Maharani et al., 2021)	IEEE	UX	30
S19	(Inan Nur et al., 2021)	Scopus	UX	61
S20	(Sinabell and Ammenwerth, 2022)	Scopus	Usability	329
S21	(Nugroho et al., 2022)	IEEE	Usability	15
S22	(Masruroh et al., 2022)	IEEE	Usability	22
S23	(Kalantari and Lethbridge, 2022)	IEEE	UX	41
S24	(Nasr and Zahabi, 2022)	IEEE	Usability	51
S25	(Saad et al., 2022)	IEEE	Usability	55
S26	(Brdnik et al., 2022)	Scopus	Both	211
S27	(Maramba et al., 2019)	Snowballing	Usability	133
S28	(Salvador et al., 2014)	Snowballing	Usability	32
S29	(Hookham and Nesbitt, 2019)	ACM	Usability	107
S30	(Prietch et al., 2022)	ACM	Both	37
S31	(Lyzara et al., 2019)	ACM	Usability	22
S32	(Lamm and Wolff, 2019)	ACM	Usability	223
S33	(Forster et al., 2018)	ACM	Both	28
S34	(da Costa et al., 2018)	ACM	Both	50
S35	(Karre et al., 2020)	ACM	Usability	36
S36	(Guerino and Valentim, 2020)	ACM	Both	39
S37	(Zhao et al., 2019)	ACM	Usability	45
S38	(Carneiro et al., 2019)	ACM	Usability	51
S39	(Böhm and Wolff, 2014)	Snowballing	Usability	55
S40	(Verkijika and De Wet, 2018)	Snowballing	Usability	18
S41	(Ren et al., 2019)	Snowballing	Usability	19
S42	(Alshamsi et al., 2016)	Snowballing	Usability	74
S43	(Petri and Wangenheim, 2017)	Snowballing	Both	117

5 RESULTS

Following the criteria, we thoroughly analyzed 43 studies to evaluate them and address our research questions systematically.

5.1 General Information

The bases selected for research were Scopus, IEEE Xplore, and ACM. Of the 43 selected studies, 13 (30.23%) accepted were found in Scopus, 13 (30.23%) in IEEE, 10 (23.26%) in ACM and 7 (16.28%) via snowballing. Papers published in conferences and journals appeared similarly: 51.16% in conferences and 48.84% in scientific journals. Regarding the scope of selected studies, 32 (74.42%) studies evaluated only primary studies focused on Usability, 5 (11.63%) evaluated only studies focused on User Experience, and 6 (13.95%) evaluated studies focused on both Usability and User Experience.

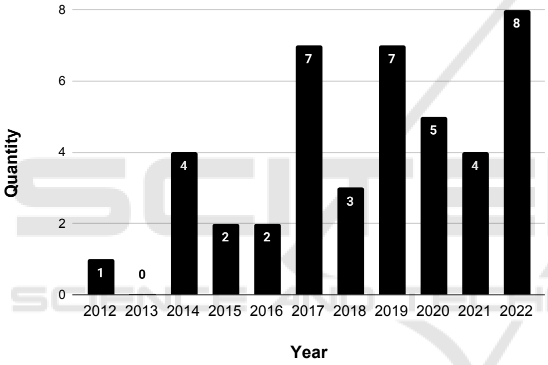

Figure 1: Publications distribution years of the 43 studies.

The timeline of publications unveiled irregularities, showcasing a distinctive inclination towards right-sided asymmetry, as illustrated in Figure 1. It initiates from a low point and experiences a substantial spike, particularly in 2017, 2019 and 2022, with 7, 7 and 9 appearances, respectively. The years 2020 and 2021 also had 5 and 4 appearances, respectively, also considered a good number. This upward trajectory aligns with the overall growth trend in publications observed in recent years. Notably, a consistent publication pattern emerged across the years, barring the absence of any studies in 2013.

The 43 studies were authored by 32 countries, as shown in Figure 2. Each study underwent an assessment wherein a coefficient was assigned, calculated as one divided by the total number of distinct author countries, with a maximum attainable value of 43. Brazil and Indonesia emerged as the frontrunners, with 5.33 appearances (12.40% of the authorship) and 5.00 appearances (11.63% of the authorship), respectively.

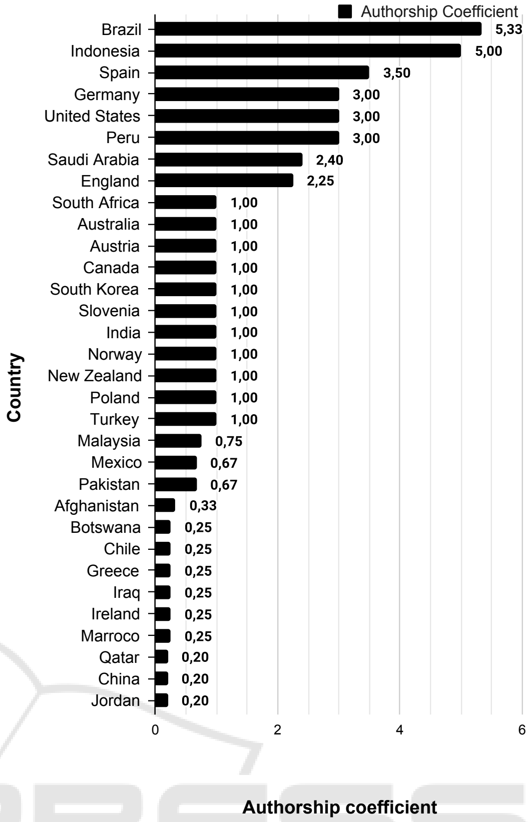

Figure 2: Countries publication distribution of the 43 studies.

Several countries attained coefficients of at least 2.00, including Spain, the United States, Peru, Germany, Saudi Arabia, and England. In contrast, 13 countries had coefficients below 1.00, indicating only partial participation in the authorship of the studies.

5.2 Applied Methods

In the context of Usability and UX, Table 4 presents the applied methods, as represented by the MRQ, in their appearances in the primary and secondary studies. The table lists the 15 main methods out of a total of 100 found, ordered by their total primary appearances, highlighting the most prominent methods identified.

To maintain consistency and clarity across multiple studies, we standardized variations in method nomenclature using singular labels. This approach streamlined the comparison of methods across different research works. However, we tried to preserve the original nomenclature used in the evaluated works, especially for methods with higher recurrence, to remain consistent with the terminology found in the studies.

Table 4: Unified methods in primary and secondary studies.

ID	Method	Primaries			Secondaries		
		Total	UX	Usab	Total	UX	Usab
M1	Questionnaire	1285	291	1160	42	11	37
M2	Usability Test	508	64	478	25	6	23
M3	Interview	387	70	342	34	9	30
M4	Observation	262	31	244	19	7	16
M5	Heuristic Evaluation	257	10	248	23	4	20
M6	Think Aloud	255	17	240	33	8	29
M7	Performance Metrics	126	30	112	13	3	11
M8	Focus Group	80	7	78	19	4	16
M9	Prototyping	76	19	75	10	4	8
M10	Experiment	66	8	58	6	1	5
M11	Cognitive Walkthrough	62	1	61	13	1	12
M12	Expert Evaluation	49	9	40	10	2	8
M13	Video Recording	29	10	22	6	2	5
M14	SUS	28	18	15	9	4	6
M15	Participatory Design	26	13	26	5	1	5

Regarding SRQ1, the Questionnaire stands out as the most used method in both Usability and UX contexts, shown in 42 of the 43 selected studies. Usability Tests, Interviews, Observations, Heuristic Evaluations, Performance Metrics, Focus groups, and Experiment methods are also significantly used. Various studies employed different names or versions of identical methods. Therefore, we unified these methods under the same label.

Furthermore, we examined the distinctions concerning usability and UX contexts. The ratio of studies focusing on Usability surpasses that of UX, approximately 3.5 times greater. Regarding citations in secondary studies, a notable dissimilarity emerges for Cognitive Walkthrough, which appears 12.0 times more frequently in the Usability context. While the positive differences for usability are minor but still evident, the Heuristic Evaluation and Experimentation methods appear proportionally 5.0 times more often.

Some methods were considerably below the average proportion of 3.5 and were, therefore, more cited in UX contexts. These were, in order, SUS, with 1.5, Prototyping, with 2.0, and Observation, with 2.3. This may demonstrate a more common recurrence of some methods in UX contexts.

Some specific methods appear expressively, leading us not to adopt their more generic classifications, *i.e.*, we present them as unique methods. This is the case of methods such as SUS and Eye Tracking, which could belong, respectively, to the Questionnaire and Sensory Measurements. Some methods were cited in a generic or unclear manner, *e.g.*, “mixed methods”, “evidence analysis” and “sampling”.

All 43 studies quantitatively addressed the methods found in their analysis of primary studies, which was also one of the objectives in the searches and definitions of the selected works. Qualitative analyses of the methods application were not evaluated in this study.

Table 5: Heatmap between studies count and methods.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15
S1	3	4	1		11	4					3	1			
S2	9		7	5		1	2								2
S3	74	83	18	5	54	31	10	7	15		6				
S4	13				7	1	2								
S5	11		14			1									
S6	104	56	41		50	38	19	6	11		11				
S7	131		29	19	2	7	18	4	1		1		6		
S8	69	6	23		17	36		9			13				
S9	10	9	3		4							6			
S10	4	7	5			2									
S11	31	10	2		28	3					2	19			
S12	13		5	2		1						4		1	
S13	15		4			1									
S14	3		2		1					18					
S15	38		1	20		2									
S16	1	1	1		1	1					1				
S17	5		5						2	22					
S18	21	2			3						1			1	
E19	46		9	3	5	6	12						7	11	
S20	4	23		1	5	1		3			3	9			
S21	16	16			5	4						1			
S22	8				5	3					6			4	2
S23	13	8	9	3	1	4		1							
S24	32	41	24			7		2							
S25	8	6	1	1				1		1				1	
S26	26	38	6			2	2	1	1	8		5			
S27	105	57	37		18	45		13							
S28	14	11			6	6	1	2	23		4				
S29	88		17	59				3							
S30	35	12	10	4		1		5	17						13
S31	2	5	2	1	11	1	5	2						2	
S32	61	77	18	63		4						2			1
S33	2													1	
S34	25		11	24		4									
S35					3			7		5	10	1			
S36	30	1	3	3	1	1	16		1				3	5	
S37	22		8	34		10		4					2		
S38	42	6	23			2	23	6				1	10		
S39	19		21	8		12	10								9
E40	2	5			11										
S41	16		9	1		3					1				
S42	13	3	6		8	10	6	4	5	12			1	2	
S43	101	21	12	6											

An analysis was also executed of which studies cite which methods. This analysis is demonstrated in Table 5. The identification of elements occurs through previous enumerations, in which “S” represents a certain study from Table 3. “M” represents a certain method from the Table 4, followed by its identifier. Three studies, S26, S33, and S34, differentiated between usability and user experience methods; however, in this table, they were combined to facilitate understanding, despite being differentiated in Table 4.

In Table 5, it is also possible to see that some studies cite many more methods compared to others, highlighting the difference in approach of each study. It is also possible to identify the continuous prominence of some methods, mainly for the Questionnaire (M1), which, in addition to being highly cited, is also generally the leader in citations for its respective study.

5.3 Primary Studies

To answer SRQ2, regarding the number of primary studies analyzed, 3021 primary studies were identified as present in secondary studies. This resulted in an average of 70.26 and a median of 45 per secondary study. Of these, the secondary study that covered fewer primary studies got 12, and the one that covered more got 329. Due to the discrepancy between the highest values, the mean is expected to be higher than the median. However, using the median, we have a more faithful average value.

The Usability context concentrated most primary studies: 2264 (74.94%). Next, we have the context for the studies that analyzed both Usability and UX, which concentrated 480 (15.89%). Lastly, there was the UX context, with 277 (9.17%). With this, it can be stated that 90.83% of the studies addressed Usability, while 25.06% addressed UX.

5.4 Evaluation Methods Classification

To assess the degree of depth and understanding of the methods analyzed and allow us to distinguish the different notions and strategies for classifying evaluation methods, SRQ3 was elaborated. It is possible to infer that most accepted studies, 29 (67.44%), do not categorize their methods, as seen in the Table 6. This may demonstrate a lack of depth and criteria for better analysis of the studies, allowing each perception and understanding of the application of the methods to be defined more concretely. A total of 9 different categorizations were found.

Table 6: Methods categorization of the 43 studies.

Category	Qty	Studies
Observation, inspection, and inquiry	5	S1;S7;S28; S31;S42
Self-report, observation, and psychophysiological	2	S19;S22
Hedonic and pragmatic	1	S12
Inspection and empirical	1	S28
Population, intervention, results, and context	1	S3
Questionnaire and interviews, inspection, and testing methods	1	S16
Expert-Recommended, Potentially Helpful, and Not Expert-Recommended	1	S20
Requirements, prototype, implementation, and mixed	1	S8
Supervised, semi-supervised, reinforcement learning, and unsupervised	1	S26
No categories mentioned	29	Remaining

Two categories are used more than one time. “Observation, Inspection, and Inquiry” was mentioned 5 times, one of which was also composed of the terms “Analytical Modeling and Simulation”. “Self-Report, Observation and Psychophysiological” had two mentions. The rest of the categories that the studies characterize are also only used in their studies, most of which are not categorizations previously found in the literature on evaluation methods.

5.5 Domains and Subdomains

To answer the question SRQ4, the domain is understood as the broad context of the study, while the subdomain is the area applied in a specific way if it exists in the scope. With this, domains and subdomains covered in the selected studies were identified.

Table 7: Domains and subdomains found in the studies.

Domain	Qty	Subdomain	Qty
Technology	35	eHealth Systems	4
Healthcare	10	Eletronic Government	3
General	6	Serious Games	3
Education	4	Healthcare	1
Accessibility	3	Chatbots	1
Governance	3	Clinical Guidelines	1
Entertainment	2	Tangible Interfaces	1
		Location-Based Games	1
		Indoor Navigation	1
		Automated Sign Language Processing	1
		Mobile Tracking	1
		Augmented Reality	1
		Virtual Reality	1
		Electronic Health Records	1
		Collaborative Health Systems	1
		Diabetes System	1
		Automated Driving System	1
		Ticket Reservation Systems	1
		Mental Health Systems	1
		University Systems	1
		Vehicle Systems	1

According to the Table 7, 35 works were identified in the Technology domain, 10 Health, 4 Education, 3 Accessibility, 3 Governance, 2 Entertainment and 6 remained in general scopes. It’s important to acknowledge that a single work can encompass multiple domains. Conversely, it’s worth mentioning that 10 studies, equivalent to 23.26% of the total, did not specify subdomains.

Of those who specified the domain, two levels of specification were defined. Considering only the most specific level, it is worth highlighting the 4 works focused on Serious Games, 4 on eHealth Systems, and 3 on Electronic Governments, respectively, from the Education, Health, and Governance domains. The other subdomains only have one appearance each. Some studies had generalist scopes, such as Mobile Applications and Software Development, not considered as specific subdomains.

5.6 Accessibility as an Evaluation Criteria

Recognizing that Accessibility is also important in HCI, it was considered, through SRQ5, to verify accessibility as a relevant factor in the reviews. As a result, it was found that 3 of the 43 studies focused on accessibility, giving it explicit attention. Another 2 considered as one of the evaluations to be made in the analysis, either as a research question or another way of being considered as a criterion to analyze.

About the 3 studies that considered accessibility a central theme, the first work (Masruroh et al., 2022) focused on considering the impact on people with general disabilities. The Questionnaire, Cognitive Walkthrough, Heuristic Evaluation, Thinking Aloud, and SUS methods were identified as the most suitable for this domain.

The second work (Nasr and Zahabi, 2022) addressed people with visual, physical, cognitive, hearing, or elderly disabilities in the indoor navigation applications scenario. The Usability Test, Questionnaire, Interview, and Think Aloud methods were found to be common to all. The third work (Prietch et al., 2022) evaluated the situation of deaf people through an analysis focused on the automatic processing of sign languages within an analysis of cultural and collaborative aspects. Other 2 studies did not focus on accessibility but considered it a topic to be evaluated in the study. One of them considered a usability evaluation of government websites and applications in Sub-Saharan Africa, in which one of the points was to evaluate accessibility (Verkijika and De Wet, 2018). The second one evaluated the usability quality of university websites in general. One of the research questions investigated the frequency of use of the term ”accessibility” within works. In it, it was seen that half of the 24 primary studies cited the term, but only 4 examined it with greater analysis (Yerlikaya and Onay Durdu, 2017). Other 8 studies had more superficial citations and considerations, citing the term and mentioning that it was important, but without an evaluation that went deeper.

6 FINDINGS AND REFLECTIONS

Based on the results presented in Section 5, some findings and discussions can be made, including a reflection on the research limitations.

The distribution of authors across countries revealed intriguing disparities. Brazil and Indonesia emerged as the primary contributors, with authorship coefficients of 5.33 and 5.00, respectively, out of a possible maximum of 43. Additionally, four more countries obtained coefficients of 3 or more in authorship: Spain, Germany, the United States, and Peru. Collectively, these six nations represented over 53% of the total authors of the 43 studies evaluated. It is important to acknowledge that this analysis may have limitations in countries where studies conducted in languages other than English predominate. This could introduce biases into the study due to the challenges of assessing non-English publications.

An extensive spectrum of methods was unearthed, culminating in nearly 100 distinct methodologies. However, most of these methods received singular mentions or were mentioned in low proportions. Notably, despite variations in applications and nomenclature, certain methods were unified due to a lack of consensus among authors.

About SRQ1, the Questionnaire method emerged as the most cited, significantly surpassing the Usability Test, which ranked as the second most cited method. These highlighted methods are entrenched in the realm of Human-Computer Interaction (HCI), underscoring their widespread recognition, applicability, cost-effectiveness, and efficiency. The prevalence of the Questionnaire method highlights its attributes of low cost, time efficiency, and simplicity. Conversely, methods relying on sensory measurements, notably Eye Tracking, were not extensively represented among the top 15 methods, potentially indicating difficulties in their application.

Regarding SRQ2, there is a noticeable difference in the number of primary studies analyzed across reviews. The smallest coverage of primary studies was observed in Araujo et al., 2014, with 12 studies, while the largest sample was that of Sinabell and Ammenwerth, 2022, with 329, interestingly both in a similar research scope. These disparities can be explained by the difference in depth between studies and the different number of studies existing on each research topic, which cannot be assessed in depth. However, the median number of studies proved to be a coherent metric, aligning well with a comprehensive review scope. The divisions of primary studies about usability and user experience maintained the proportionality of the previously mentioned data.

In SRQ3, an absence of standardization in categorizing methods was evident. The most common categorization was found in 5 of the 43 studies, 11.63% of them. Furthermore, 67.44% chose not to declare any categorization for their methods. This lack of categorization, especially in studies not centered on computing, possibly indicates a lack of knowledge or perceived necessity for in-depth classification.

To discuss the SRQ4, technological domains were prominent, with approximately 81.40% of studies encompassing some technological definition, as can be seen in Table 7. Notably, Health emerged as a crucial domain, with 10 recurrences (23.26%), signifying a significant focus on evaluating usability and user experience within healthcare solutions. Subdomains like eHealth Systems, Electronic Government, and Serious Games garnered multiple appearances, highlighting diverse thematic applications.

Referring to SRQ5, the consideration of Accessibility as an evaluation criterion within the 43 studies was relatively limited. Only 3 studies directly addressed accessibility, while 2 studies evaluated it at some point in their work. Given the empathetic scope of this study towards users in computer systems, this relatively low attention to accessibility poses a potential gap warranting further investigation.

Several potential limitations were identified, impacting the validity of this research. Factors such as overlooking primary study information within secondary studies – as the year of application or verify if another secondary study has already cited the primary study –, exclusion of ”UX” in the search string, limited number of databases, and a lack of in-depth analysis of secondary study quality serve as limitations to be acknowledged and addressed in future studies.

The application of methods to evaluate usability and user experience has been the subject of study in HCI literature. This was possible to identify through an initial exploratory analysis, inspecting the number of secondary studies, obtaining 27 secondary studies between 2012 and 2021. However, there is a lack of having an analytical tertiary study on the topic, due to the number of existing secondary studies and the investigative possibilities that such a study would bring, defined by the research questions of this work.

Following the guidelines of Petersen et al., 2008, and Petersen et al., 2015, a systematic mapping of the literature on secondary studies related to the topic was carried out. Research questions, exclusion, and inclusion criteria were developed, in addition to the analysis of the selected secondary studies.

During data extraction, it was possible to answer all research questions. A predominance of studies addressing usability in comparison to user experience was noted. A preference for more widespread methods that are also easier to apply was also identified, mainly a predominance of the Questionnaire method. A good diversity was observed between the countries that authored the studies, with a slightly greater dominance of Brazil and Indonesia, as well as in the themes covered by the studies, with emphasis on works in the health domain, present in 10 of the 43 studies. An increasing trend in the publication of related studies was also noted, obtaining more results in more recent years compared to previous years.

Another point of analysis was the issue of considering accessibility as an evaluation classification in studies. Very few studies were identified considering accessibility as a direction to be evaluated in secondary studies. Only 2 of the 43 studies made this consideration. 3 works had accessibility as the central scope of the work, each with its target group conception. However, this may suggest that accessibility is only considered when the main topic of the study, but is somewhat taken into account when a relevant attribute is related to usability or user experience. This result may suggest a fragmentation in the understanding of the relations between the concepts of usability and user experience with accessibility, opening space for more detailed investigations.

The obtained results yielded pertinent findings that advance future research intentions in the field. The identification of the most utilized methods is deemed valuable. Other information, such as methods categorization, methodologies used, countries and years of publication of studies and current accessibility considerations in studies, can help to understand a little better the current situation of secondary studies within the topic of this research. As well as the threats found and results that suggest gaps and uncertainties for further investigation. This allows the study to also serve as a basis and encourage future studies.

ACKNOWLEDGEMENTS

This research is partially supported by CNPq grant 302959/2023-8 and 308395/2020-4 (DT2), FAPESC Edital nº 48/2022 TO n°2023TR000245 and CAPES - Financing Code 001.

REFERENCES

ABNT (2011). *ABNT NBR ISO/IEC 9241 - Ergonomia da interação humano-sistema - Parte 210: Projeto centrado no ser humano para sistemas interativos*. Associação Brasileira de Normas Técnicas - ABNT NBR.

Almazroi, A. A. (2021). A systematic mapping study of software usability studies. *International Journal of Advanced Computer Science and Applications*, 12(9).

Alshamsi, A., Williams, N., and Andras, P. (2016). The trade-off between usability and security in the context of egovernment: A mapping study. In *Proceedings of the 30th International BCS Human Computer Interaction Conference (HCI)*.

Ansaar, M. Z., Hussain, J., Bang, J., Lee, S., Shin, K. Y., and Young Woo, K. (2020). The mhealth applications usability evaluation review. In *2020 International Conference on Information Networking (ICOIN)*, pages 70–73.

Araujo, L. P. d., Berkenbrock, C. D. M., and Mattos, M. M. (2014). A systematic literature review of evaluation methods for health collaborative systems. In *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 366–369.

Barbosa, S. D. J., Silva, B. S. d., Silveira, M. S., Gasparini, I., Darin, T., and Barbosa, G. D. J. (2021). Interação humano-computador e experiência do usuario. *Auto publicação*.

Böhm, V. and Wolff, C. (2014). A review of empirical intercultural usability studies. pages 14–24, Cham. Springer International Publishing.

Brdnik, S., Heričko, T., and Šumak, B. (2022). Intelligent user interfaces and their evaluation: A systematic mapping study. *Sensors*, 22(15).

Buchinger, D., Cavalcanti, G., and Hounsell, M. (2014). Mecanismos de busca acadêmica: uma análise quantitativa. *Revista Brasileira de Computação Aplicada*, 6(1):108–120.

Buse, R. P., Sadowski, C., and Weimer, W. (2011). Benefits and barriers of user evaluation in software engineering research. In *Proceedings of the 2011 ACM International Conference on Object Oriented Programming Systems Languages and Applications*, OOPSLA '11, page 643–656, New York, NY, USA. Association for Computing Machinery.

Carneiro, N., Darin, T., and Viana, W. (2019). What are we talking about when we talk about location-based games evaluation? a systematic mapping study. IHC '19, New York, NY, USA. Association for Computing Machinery.

da Costa, V. K., de Vasconcellos, A. P. V. a., Darley, N. T., and Tavares, T. A. (2018). Methodologies and evaluation tools used in tangible user interfaces: A systematic literature review. IHC 2018, New York, NY, USA. Association for Computing Machinery.

da Silva Osorio, A. F., Schmidt, C. P., and Duarte, R. E. (2008). Parceria universidade-empresa para inclusão digital. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*, pages 308–311.

Ellsworth, M. A., Dziadzko, M., O'Horo, J. C., Farrell, A. M., Zhang, J., and Herasevich, V. (2017). An appraisal of published usability evaluations of electronic health records via systematic review. *Journal of the American Medical Informatics Association*, 24(1):218–226.

Feather, J. S., Howson, M., Ritchie, L., Carter, P. D., Parry, D. T., and Koziol-McLain, J. (2016). Evaluation methods for assessing users' psychological experiences of web-based psychosocial interventions: A systematic review. *Journal of medical Internet research*, 18(6):e5455.

Fernandez, A., Abrahão, S., and Insfran, E. (2012). A systematic review on the effectiveness of web usability evaluation methods. In *16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012)*, pages 52–56.

Forster, Y., Hergeth, S., Naujoks, F., and Krems, J. F. (2018). How usability can save the day - methodological considerations for making automated driving a success story. AutomotiveUI '18, page 278–290, New York, NY, USA. Association for Computing Machinery.

Guerino, G. C. and Valentim, N. M. C. (2020). Usability and user experience evaluation of conversational systems: A systematic mapping study. SBES '20, page 427–436, New York, NY, USA. Association for Computing Machinery.

Hookham, G. and Nesbitt, K. (2019). A systematic review of the definition and measurement of engagement in serious games. ACSW '19, New York, NY, USA. Association for Computing Machinery.

Inan Nur, A., B. Santoso, H., and O. Hadi Putra, P. (2021). The method and metric of user experience evaluation: A systematic literature review. ICSCA 2021, page 307–317, New York, NY, USA. Association for Computing Machinery.

Kalantari, R. and Lethbridge, T. C. (2022). Characterizing ux evaluation in software modeling tools: A literature review. *IEEE Access*, 10:131509–131527.

Karre, S. A., Mathur, N., and Reddy, Y. R. (2020). Understanding usability evaluation setup for vr products in industry: A review study. *SIGAPP Appl. Comput. Rev.*, 19(4):17–27.

Khodambashi, S. and Nytrø, Ø. (2017). Usability methods and evaluation criteria for published clinical guidelines on the web: A systematic literature review. In *International Conference on Human-Computer Interaction*, pages 50–56. Springer.

Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

Lamm, L. and Wolff, C. (2019). Exploratory analysis of the research literature on evaluation of in-vehicle systems. AutomotiveUI '19, page 60–69, New York, NY, USA. Association for Computing Machinery.

Lyzara, R., Purwandari, B., Zulfikar, M. F., Santoso, H. B., and Solichah, I. (2019). E-government usability evaluation: Insights from a systematic literature review. ICSIM 2019, page 249–253, New York, NY, USA. Association for Computing Machinery.

Maharani, L., Durachman, Y., and Ratnawati, S. (2021). Systematic literature review method for evaluation of user experience on ticket booking applications. In *2021 9th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–7.

Maia, C. L. B. and Furtado, E. S. (2016). A systematic review about user experience evaluation. volume 9746, pages 445–455, Cham. Springer International Publishing.

Maramba, I., Chatterjee, A., and Newman, C. (2019). Methods of usability testing in the development of ehealth applications: A scoping review. *International Journal of Medical Informatics*, 126:95–104.

Masruroh, S. U., Rizqy Vitalaya, N. A., Sukmana, H. T., Subchi, I., Khairani, D., and Durachman, Y. (2022). Evaluation of usability and accessibility of mobile application for people with disability: Systematic literature review. In *2022 International Conference on Science and Technology (ICOSTECH)*, pages 1–7.

Nagalingam, V. and Ibrahim, R. (2015). User experience of educational games: a review of the elements. *Procedia Computer Science*, 72:423–433.

Nasr, V. and Zahabi, M. (2022). Usability evaluation methods of indoor navigation apps for people with disabilities: A scoping review. In *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, pages 1–6.

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Norman, D. (2014). *Things that make us smart: Defending human attributes in the age of the machine*. Diversion Books.

Nugroho, A., Santosa, P. I., and Hartanto, R. (2022). Usability evaluation methods of mobile applications: A systematic literature review. In *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)*, pages 92–95.

Paz, F. and Pow-Sang, J. A. (2014). Current trends in usability evaluation methods: A systematic review. In *2014 7th International Conference on Advanced Software Engineering and Its Applications*, pages 11–15.

Paz, F. and Pow-Sang, J. A. (2015). Usability evaluation methods for software development: A systematic mapping review. In *2015 8th International Conference on Advanced Software Engineering & Its Applications (ASEA)*, pages 1–4.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, pages 1–10.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.

Petri, G. and Wangenheim, C. G. v. (2017). How games for computing education are evaluated? a systematic literature review. *Comput. Educ.*, 107(C):68–90.

Prietch, S., Sánchez, J. A., and Guerrero, J. (2022). A systematic review of user studies as a basis for the design of systems for automatic sign language processing. *ACM Trans. Access. Comput.*, 15(4).

Ren, R., Castro, J., Acuña, S., and Lara, J. (2019). Usability of chatbots: A systematic mapping study. pages 479–484.

Rogers, Y., Sharp, H., and Preece, J. (2013). *Design de interação*. Bookman Editora.

Saad, M., Zia, A., Raza, M., Kundi, M., and Haleem, M. (2022). A comprehensive analysis of healthcare websites usability features, testing techniques and issues. *IEEE Access*, 10:97701–97718.

Saare, M. A., Hussain, A., Jasim, O. M., and Mahdi, A. A. (2020). Usability evaluation of mobile tracking applications: A systematic review. *Int. J. Interact. Mob. Technol.*, 14(5):119–128.

Salvador, C., Nakasone, A., and Pow-Sang, J. A. (2014). A systematic review of usability techniques in agile methodologies. EATIS '14, New York, NY, USA. Association for Computing Machinery.

Sheikh, S., Bin Heyat, M. B., AlShorman, O., Masadeh, M., and Alkahatni, F. (2021). A review of usability evaluation techniques for augmented reality systems in education. In *2021 Innovation and New Trends in Engineering, Science and Technology Education Conference (IETSEC)*, pages 1–6.

Sinabell, I. and Ammenwerth, E. (2022). Agile, easily applicable, and useful ehealth usability evaluations: Systematic review and expert-validation. *Applied clinical informatics*, 13(01):67–79.

Soares, M. M., Rebelo, F., and Ahram, T. Z. (2022). Handbook of usability and user-experience: Research and case studies. volume 1. CRC Press.

Verkijika, S. F. and De Wet, L. (2018). A usability assessment of e-government websites in sub-saharan africa. *International Journal of Information Management*, 39:20–29.

Weichbroth, P. (2020). Usability of mobile applications: A systematic literature study. *IEEE Access*, 8:55563–55577.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, EASE '14, New York, NY, USA. Association for Computing Machinery.

Yanez-Gomez, R., Cascado-Caballero, D., and Sevillano, J.-L. (2017). Academic methods for usability evaluation of serious games: a systematic review. *Multimedia Tools and Applications*, 76(4):5755–5784.

Yerlikaya, Z. and Onay Durdu, P. (2017). Usability of university websites: a systematic review. In *International Conference on Universal Access in Human-Computer Interaction*, pages 277–287. Springer.

Zapata, B. C., Fernández-Alemán, J. L., Idri, A., and Toval, A. (2015). Empirical studies on usability of mhealth apps: a systematic literature review. *Journal of medical systems*, 39(2):1–19.

Zarour, M. and Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. *Cogent Engineering*, 4(1):1421006.

Zhao, L., Loucopoulos, P., Kavakli, E., and Letsholo, K. J. (2019). User studies on end-user service composition: A literature review and a design framework. *ACM Trans. Web*, 13(3).