A Systematic Mapping Study of Software Usability Studies

Abdulwahab Ali Almazroi
University of Jeddah, College of Computing and Information Technology at Khulais
Department of Information Technology, Jeddah, Saudi Arabia

***Abstract*—Among software quality attributes "software usability" is considered as one of the vital factors in software engineering literature. Software usability is the ability for users to generally understand, use, and learn a software with ease. Due to the importance of usability in software quality, a considerable amount of literature is published in the past decade. Few review and survey studies are also published to critically review the existing literature in the domain. However, there is limited research covering systematic mapping study of software usability. Mapping studies help in analyzing the general trends and research productivity in a research area. To fill this gap, this work critically examines the overall research productivity, demographics, trends, and challenges of software usability. The objective is to classify the current contributions and trends in the area of software usability. We retrieved 9,874 research articles from six research databases and 62 works are selected as primary studies using an evidence-based approach. The result of this mapping study shows that software usability is an active research area, with a promising number of works published in the last decade (2011 - 2020). We identified that the current literature spans over multiple article classes of which investigative papers, model proposals and evaluation papers are the most frequently published article types. We found experiments and theoretical validations to be the most common validation techniques. In terms of application domains; web, software development and mobile applications are the most frequent domains where usability studies are conducted. We identified that future usability studies should focus more on field studies as well as on the usability testing of scientific software packages. It will be of importance to consider ethical issues in usability testing as well.**

Keywords*—Software usability; usability study; systematic mapping study; systematic literature review; software engineering

I. Introduction

Based on IEEE Std.610.12, software usability is the user's simplicity in learning to provide inputs and operate a given component or system [1]. Although software and system usability are considered as non-functional requirements, its importance cannot be overstated [2]. Sagar and Saha [3] listed six software quality criteria which includes usability as well. As modern-day software is evolving to be more complex and omnipresent, software usability has become an indispensable non-functional requirement for ensuring software quality.

In the field of software engineering, various studies have identified usability issues and its implication on software quality [4-6]. As per ISO/IEC 9126, software usability is the ability of users to generally understand, use, and learn the software with ease. Based on [7], there are five important usability attributes, which are learnability, attractiveness, understandability, operability, and usability compliance. Thus, the definition for software usability differs among different standards and researchers [3].

In the real world, various application domains have duly considered usability engineering as an important area. It is widely considered to be a far-reaching research area at large [3] and is used in domains such as aerospace [8], computer-based medical devices [9], defense[10], mobile devices [11], web applications [12] etc. Various research and review studies were conducted on software usability [3, 11-21]. However, based on our findings from the existing literature, we observed that a systematic mapping study (SMS) in this domain is lacking, confirming the claim of Bitkina, et al. [9] regarding the absence of UX/usability studies. Hence, this SMS is undertaken to fill this research gap through the extensive analysis of important studies that were published in the last decade (2011-2020). Filling this research gap will help researchers and experts better understand software usability's efficiency, effectiveness, and development.

In an effort to fill the research gap and inclusion of significant works, this research followed systematic mapping methodology. This approach allows the research to capture key facts and details from literature using a well-defined process. To this end, an SMS protocol composed of search strategy, data extraction, selection criteria, and rejection criteria was formed. The main objective of this research is to investigate factors affecting software usability. Additionally, this researched also aimed at classifying the selected studies by knowing the existing contributions, research facets conducted, validation methods used, evaluation measures utilized, application domains, and lastly the overall demographics of the literature reviewed. The selected parameters will provide an overview of the general trends of the publications as well as the evolution of research in the domain of software usability.

In this study, the main contributions are as follows:

- A detailed examination and synthesis of key studies on software usability.
- The study reviews primary studies (PS) and identify their distinct contributions.
- The mapping study analyzes the overall research productivity, demographics, trends, and challenges in software usability.

- The research identifies area of research that are least addressed and provides directions for future research.

This research study is divided into six sections. Section II presents various aspects of software usability and discusses related surveys in the domain of software usability. Section III describes the research methodology including research questions, data acquisition and processing techniques. Results are presented in Section IV. Section V discusses the research findings as well as directions for future research, whereas threats to validity are discussed in Section VI. Section VII concludes the work.

II. Software Usability and Related Work

This section discusses software usability by reviewing articles published in the area.

A. Software Usability

Software usability is a key characteristic of software quality. As per ISO/IEC 9126, usability is defined as "a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users" [3, 7]. Based on [7], software quality model is composed of six dimensions, which are functionality, portability, maintainability, efficiency, usability and reliability.

However, usability is inconsistently defined in the literature, where different standards and researchers have distinct definitions. In the following text, some definitions of software usability from various standards are outlined. In a study by Nielsen, the author defined usability based on five key dimensions. These dimensions are efficiency, errors, learnability, memorability, and satisfaction [22]. Efficiency means that the system should be efficient to be utilized by a user. When the system is efficiently utilized, productivity will increase. It is also expected that such systems have low error rates. Learnability is the ability to understand the system, and a good system is expected to have small learning curve. Memorability refers to the ability of the system to be easily remembered without requiring users to relearn things again. Lastly, satisfaction means that the user should feel gratified when using the system. Moreover, in the ISO 9241-11, usability is defined as the level where a product can be utilized by a given user in achieving a specific objective with efficiency, effectiveness, and satisfaction in a defined context [23, 24].

In software engineering, usability issues in software directly or indirectly contribute to software quality problems. These problems often cause low efficiency and effectiveness. Based on this, users encounter various difficulties when using a specific software [25]. Further, the usability issues can also lead to low acceptance rate of software applications [26].

B. Related Work

This section summarizes the existing literature in software usability and identifies the research gaps.

Bastien [27] focused on the growing applications of Information and Communication Technology (ICT) based medical devices incorporating human-computer interfaces. The author highlighted that usability of these medical devices' hinges on their ease of usability. The author identified inspection-, user- and model-based evaluations as standard industry approaches for validating usability. The work considered various scenarios and provided directions for further research in usability testing for mobile applications. Usability testing can be performed in a number of manners such as questionnaires, interviews, automated testing etc. Sure [21] considered questionnaire as a method for usability testing where information about usability, user satisfaction, expectations, software behavior and other related information is collected using a set of questions. The author identified thirteen questionnaires from literature. The analysis of the questionnaires revealed that 'satisfaction' is the most common factor that is measured by the majority of the questionnaires.

Esmeria and Seva [18] presented a survey of works in relation to the usability of the websites. The authors argued that the process of usability testing should result in a usability index that can describe the usability of a website. A drawback of the work is the lack of rigor in selection of data sources. The authors relied on Science Direct and ACM Digital Library only and no specific query string was mentioned. Sagar and Saha [3] conducted a review of the software usability and reached to a variety of conclusions. The authors identified that so far researchers have not been able to unanimously agree on usability models. Further, the work highlighted efficiency, effectiveness, satisfaction, and learnability as standard measures used by a majority of the usability models. In terms of usability evaluation methods, questionnaires, usability testing and heuristic evaluations are the leading mechanisms.

Maramba et al. [28] conducted a systematic review of usability testing of e-health applications. The authors observed that although there is an exponential increase in the development and subsequence deployment of e-health applications, very few of these applications have published their usability evaluation results. The work further highlighted that questionnaire has been the de-facto method for usability analysis and argued that more qualitative and automated methods must be used in usability evaluation. Key limitations of the work include the limited time frame for the selection of potential works (April 2014 – October 2017) and the scope of the work which is limited to e-health applications only. Weichbroth [29] used Scopus as data source to conduct a systematic literature review of usability aspects of mobile applications. The author identified 790 relevant documents spanning 2001 to 2018. It was observed that the usability definition as given in ISO 9241-11 is used by majority of the works. The work identified 75 attributes associated with the usability of mobile applications and identified efficiency, satisfaction and effectiveness as most important considerations as indicated in the literature. Memorability, cognitive load and errors are identified as least considered attributes. Like Maramba, et al. [28], and Weichbroth [29] observed that controlled observations and surveys are mostly used in usability studies and suggested the use of eye-tracking, thinking loud, and interviews to be used in the future usability studies. Besides these studies, other important works include Coursaris and Kim [30], Harrison, et al. [31], and Quiñones and Rusu [32].

TABLE I. SUMMARY OF THE SELECTED SURVEYS

S. No	Study and Year	Selected Databases	Keywords	No. of Selected Articles	Limitations
1	Harrison, et al. [31] (2013)	ACM Digital Library, Google Scholar, IEEE Xplore	Mobile application evaluations, mobile application usability evaluations, usability of mobile applications	131	Narrow scope focusing on mobile applications only Unavailability of 6% of the selected papers
2	Sure [21] (2014)	ACM Digital Library, Inspec	Usability, Questionnaire, Reliability, Validity	35	Generic query terms Use of Inspec and ACM Digital library for literature search Limited scope focusing on questionnaires for a usability study
3	Esmeria and Seva [18] (2017)	ACM Digital Library, ScienceDirect	Website usability evaluation, measures of web usability	42	Lack of rigor in selection of data sources Reliance on ACM Digital Library and Science Direct only
4	Quiñones and Rusu [32] (2017).	ACM Digital Library, Science Direct, IEEE Xplore, Springer Link, Scopus, Google Scholar	Usability heuristic(s), methodology, heuristic evaluation, formal process, usability design	73	Focused on usability heuristics only
5	Maramba, et al. [28] (2019)	ACM Digital Library, CINAHL, IEEE Xplore, Medline / PubMed.	eHealth, mHealth, usability	133	Limited focus on e-health applications only Very short time frame selection
6	Weichbroth [29] (2020)	Scopus	Usability, mobile applications	66	A narrow focus on mobile applications Use of single source for data collection

Table I presents a summary of the selected related works. Note that some works such as Bastien [27] and Coursaris and Kim [30] are not included in the table as these works did not use the standard methodology for conducting systematic literature review.

Despite the usefulness of these studies, none of the highlighted studies in this section conducted a systematic mapping study for software usability. Most papers study selection process is also arbitrary with no rigor or repeatability. Hence, we observed that there is no study in the software usability that categorize and analyze existing research with respect to their research facets, contribution facets, publication forums/trends, citation impacts, and so on. Thus, the aim of this research is to fill these gaps in the field.

III. METHODOLOGY

Systematic mapping studies (SMS) are conducted to provide a general overview of a research area by systematically classifying the existing works and identify the contributions of researchers in the area of study. The studies (SMS) largely explore current literature to examine the reporting of areas, publication frequency, research trends, and publication venues where the primary studies are published [33, 34]. There are many shared characteristics between a systematic literature review (SLR) and systematic mapping studies. Some of the characteristics are the use of evidence-based searching and study selection procedures. However, SMS has a distinct objective from SLR and took a different approach to data analysis. SMS is primarily aimed at mapping and structuring an area of study. Hence, in this study, the SMS follows the general guidelines suggested by Petersen, et al. [34] and Kitchenham and Brereton [35]. Consequently, this study follows the pathway of similar studies that adopt these guidelines [36-38]. Fig. 1 outlines the process for our study.

Fig. 1. The Systematic Mapping Process.

As depicted in Fig. 1, the SMS process comprises of five distinct stages. The initial stage is to define key research questions (RQs). In stage two, the search process is conducted by specifying the search terms for retrieving the primary studies (PS). In the third and fourth stages, the retrieved studies are screened to remove unnecessary and irrelevant studies. Lastly, the data extraction is conducted, and the systematic maps of the study are created. These stages are defined in the following text.

A. Research Questions

In this section, the research questions (RQs) of the study are presented with respect to our main research objective. The main objective of this SMS is to classify the selected studies by knowing the existing contributions, research facets, validation methods, evaluation measures, application domains, and the overall demographics of the selected studies. The key question of this SMS is "what is the state-of-the-art in software usability studies?". Based on the objective of the work, the key question is divided into six distinct RQs as presented in Table II.

B. Data Sources

A SMS heavily relies on the selected Primary Studies (PS). To achieve a good mapping, it is vital that the process of selecting the PS is conducted carefully. The literature on software usability was collected from 2011 to 2020. Six data sources were selected for our literature search. The data sources include ACM Digital Library, IEEE Xplore, Google Scholar, Science Direct, SpringerLink, and Taylor and Francis.

It is important to mention that the selected sources include most the publications in the area of usability. Scopus could not be included as the author was unable to obtain access to the restricted (subscription based) database of Scopus. However, the selected six sources for data retrieval are broad and comprehensive enough to provide coverage to most of the reputable publications' outlets. In addition, the selected sources are used in other studies such as Esmeria and Seva [18] and Maramba, et al. [28]. Table III highlights the data sources with respect to the studies identified in our initial result search. The initial search resulted in 9,874 studies, out of which 62 studies were shortlisted based selection criteria covered in Section 3.D.

TABLE II. THE DEFINED RESEARCH QUESTIONS

RQ#	Research Question	Motivation
RQ1	What are the demographic characteristics of the PS?	To identify publication trend, publication forums and citation impact of the primary studies.
RQ2	What contribution facets have the primary studies provided?	To identify the contribution facets (model, method, investigation, and so on).
RQ3	What are the research types (facet) focused on by the primary studies in the domain?	To ascertain the research facets (evaluation research, solution proposal, and so on) in the area of study.
RQ4	What validation methods are generally utilized for software usability evaluation?	To identify various software usability evaluation methods.
RQ5	What are the various application domains for usability?	To investigate the domains of application for usability.
RQ6	What are the evaluation measures used by the PS?	To identify the evaluation measures used by the selected studies.

TABLE III. STUDIES IDENTIFIED IN EACH DATA SOURCES

Data Source		Initial Results of Search	Final Selected Studies
ID	*Name*		
D1	IEEE Xplore	729	18
D2	Science Direct	1,208	15
D3	Taylor and Francis	76	12
D4	ACM	957	7
D5	SpringerLink	467	5
D6	Google Scholar	6,437	5
Total		9,874	62

C. Search Terms

We performed automatic searches to retrieve important studies from our selected data sources. This is achieved using our search string or terms developed based on the guidelines of Petersen, et al. [34]. Basically, a search string is a composition of characters used by a researcher to identify the most relevant set of documents from a data source. Therefore, selecting the right search string is imperative because the outcome is connected to the information given by the data source. Hence, the selection of right terms requires careful attention to ensure important studies are not missed in the mapping process. In doing so, a generic search string was formulated to work on all the data sources. The search string is outlined as follows.

((Software usability AND Usability models) OR (Usability metrics))

It is important to point out that the search term is relatively generic to ensure that maximum results are obtained. A carefully crafted inclusion and exclusion criterion is then applied to shortlist the publications for further study. After execution of the search terms on respective data sources, the resultant publications obtained from each of the data source are reflected in Table III. A total of 9,874 publications were retrieved. Note that the search query was executed to collect the raw data from 19 November 2020 to 25 December 2020.

D. Inclusion and Exclusion Criterion

When the search results utilizing the articulated search string are acquired, we anticipated that works that are not important to the objective of this SMS study can also be retrieved. If so happens, we cautiously designed a clear inclusion and exclusion criteria that will be used on the retrieved studies to eliminate those that are not in-line with the objective of the paper. The inclusion-exclusion criteria for this study are outlined as follows;

Inclusion Criteria:

- Include studies on software usability
- Include studies that were published in the past 10 years (2011 - 2020)
- Include only peer-reviewed studies

Exclusion Criteria:

- Exclude survey and review studies
- Short Papers
- Editorials
- Summaries of keynotes
- Exclude the studies that are not on software usability
- Exclude the studies that are not written in English

After a careful implementation of inclusion and exclusion criterion, and thorough manual analysis including removal of duplicates, we identified 62 primary studies for further analysis.

E. Extraction of Data

To extract data from each of the 62 primary studies for answering the formulated RQs, a systematic data extraction method has to be clearly defined. We created a form to extract important data from the 62 identified articles for this study. The author as well as the two volunteers filled out the form for each of the 62 selected papers. For each publication, title, publication venue, research type, contributions, validation method, publication year, evaluation measures and application domain were recorded.

F. Classification Scheme

We followed Petersen, et al. [34] to develop the classification scheme for this study. The 62 final selected studies were examined by their titles, abstracts, keywords, research contributions, theoretical models, and general demographics. These studies were comprehensively studied for a thorough understanding of various characteristics of the classification. The first step is to classify the PS into the contributions made by various researchers. These contributions are investigative study, evaluation study, model, framework, application, scheme, method, usability concepts, usability principles, approach, and system. We further extended the classification to identify the research facets. These facets are experience papers, evaluation research, solution proposals, and validation research. These classifications are standard and in line with the existing literature for conducting a mapping study [34, 38]. Subsequently, further classifications of the PS were done with respect to the validation methods used to validate software usability, evaluation measures, application domains, and the selected studies general demographic characteristics.

IV. Results

This section covers the results of the research. The RQs formulated are all answered by critically analyzing the PS studies. Table IV presents the PS for this study. For the sake of brevity, only research article IDs is provided. For mapping between IDs and research articles description, the reader is referred to Appendix A.

TABLE IV. Overview of Selected Studies

PS Paper ID	Year of Publication	Publication Channel	Citation Count	Contribution
B1	2020	IEEE	2	Approach
B2	2017	ACM	5	Model
B3	2017	ACM	2	Model
B4	2019	ACM	0	Evaluation
B5	2020	Springer	1	Evaluation
B6	2019	Springer	1	Metrics
B7	2017	IEEE	4	Evaluation
B8	2017	IEEE	1	Evaluation
B9	2017	IEEE	3	Investigation
B10	2018	ACM	2	Investigation
B11	2018	IEEE	0	Investigation
B12	2018	IEEE	1	Method
B13	2016	Taylor and Francis	13	Usability concepts
B14	2015	Taylor and Francis	62	Investigation
B15	2016	Taylor and Francis	10	Investigation
B16	2016	Taylor and Francis	3	Evaluation
B17	2015	Taylor and Francis	42	Investigation
B18	2011	Elsevier	92	Investigation
B19	2013	Taylor and Francis	78	Usability principles
B20	2016	Taylor and Francis	2	Investigation
B21	2012	Elsevier	284	Investigation
B22	2013	IEEE	78	Usability guidelines
B23	2011	Elsevier	74	Evaluation
B24	2012	Elsevier	75	Investigation
B25	2013	Elsevier	63	Investigation
B26	2011	ACM	17	Model
B27	2020	IEEE	0	Evaluation
B28	2020	IEEE	1	Investigation
B29	2011	Independent	60	Investigation
B30	2013	Springer	23	Evaluation
B31	2016	Springer	103	Evaluation
B32	2014	Independent	113	Evaluation
B33	2019	IEEE	1	Evaluation
B34	2019	IEEE	0	Investigation
B35	2018	IEEE	0	Model
B36	2018	IEEE	1	Investigation
B37	2019	IEEE	1	Model
B38	2013	Elsevier	66	Framework
B39	2019	IEEE	4	System
B40	2014	Elsevier	24	Scheme
B41	2020	IEEE	0	Investigation
B42	2015	Independent	1	Approach
B43	2019	IEEE	0	Investigation
B44	2014	Elsevier	34	Evaluation
B45	2015	Taylor and Francis	4	Investigation
B46	2015	Taylor and Francis	118	Evaluation
B47	2012	Elsevier	25	Application
B48	2015	Elsevier	20	Model
B49	2012	Taylor and Francis	83	Evaluation
B50	2013	Springer	12	Evaluation
B51	2012	ACM	20	Investigation
B52	2013	Taylor and Francis	14	Investigation
B53	2013	IEEE	1	Model
B54	2011	ACM	8	Evaluation
B55	2013	Taylor and Francis	33	Evaluation
B56	2015	Elsevier	15	Application
B57	2013	Independent	3	Model
B58	2015	Elsevier	29	Investigation
B59	2013	Elsevier	8	Investigation
B60	2015	Elsevier	21	Approach
B61	2013	Elsevier	47	Evaluation
B62	2011	Independent	1	Investigation

A. RQ1. What are the Demographics Characteristics of the PS?

In answering this RQ, the primary studies were analyzed critically with the purpose of answering the RQ. Three aspects of the PS were analyzed including publication trend, publication forums, and citation impact.

Publication trend: From 2011 to 2020, 62 studies were retrieved from the data sources. In Fig. 2, the year-wise publications in the domain of software usability are graphically presented. We observed that in 2013 and 2015, more studies were published with 12 and 9 studies, which are the most active years in the research domain. 2019 was also moderately active, with 7 studies. In general, even though the number of studies is linear, the research output continues to stabilize with stable yearly publication.

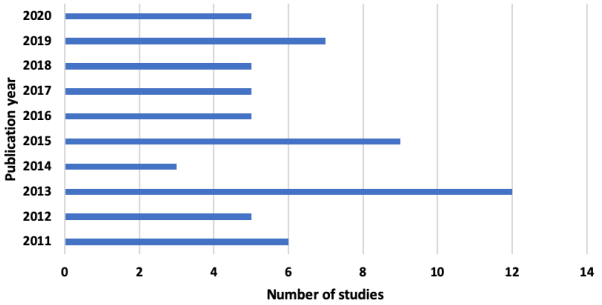

Fig. 2. Publication Trend Per Year.

Publication forums: This SMS study covered 20 different journals, 19 different conference proceedings, and 1 symposium and workshop each, respectively (see Appendix B). 39 papers are published in journals, 21 in conferences and one study each in workshop and symposium. From the analysis, we found that the International Journal of Human-Computer Interaction and the International Journal of Human-Computer Studies were the venues that contribute the most with 7 and 6 publications respectively followed by Journal of Systems and Software and Behavior and Information Technology with 4 and 2 publications each.

Citation impact: In Table V, the number of citations for the top 10 most cited papers is presented. We obtained the citation count of each paper from Google Scholar. Hence, the citation count may or will change at any point in time. In general, from our PS, we found four papers that have more than 100 citations each, which are Lee and Kozar [39], Kortum and Sorber [40], Mirkovic, et al. [41], and Maitama, et al. [36]. The total number of citations from the PS was 1809 as presented in Table IV. Therefore, the average number of citations per paper is 29.17.

B. RQ2. What Contribution Facets have the PS Provided?

To answer the RQ, we conducted a thorough analysis of the selected PS. Based on the analysis, we found 13 key contributions. These contributions are summarized in Fig. 3. The most significant contributions are Investigation (23 papers), Evaluation (18 papers), and Model (8 papers). 37% of the PS conducted an investigative study on software usability, followed by evaluation (29%), and model (13%), respectively. The rest of the contributions have less than 5% coverage. From our analysis, we observed that majority of the studies conducted an investigation into usability of existing web applications (B6, B15, B8, B38) or usability evaluation models (B37, B57, B53). Other studies focused more on proposing new models to help in facilitating or understanding key factors that facilitate or hinder usability of software or web application.

TABLE V. TOP CITED PAPERS

PS Paper ID	Paper Title	Citation	Year
B21	Understanding of website usability: Specifying and measuring constructs and their relationships	284	2012
B46	Measuring the usability of mobile applications for phones and tablets	118	2015
B32	Supporting cancer patients in illness management: usability evaluation of a mobile app	113	2014
B31	Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards	103	2016
B18	Reliability, validity, and sensitivity of a single-item measure of online store usability	92	2011
B49	A comparison of usability evaluation methods for evaluating e-commerce websites	83	2012
B19	Usability principles for augmented reality applications in a smartphone environment	78	2013
B22	Usability through software design	78	2013
B23	Aesthetics and usability of in-vehicle navigation displays	74	2011
B24	How do usability professionals construe usability?	75	2012

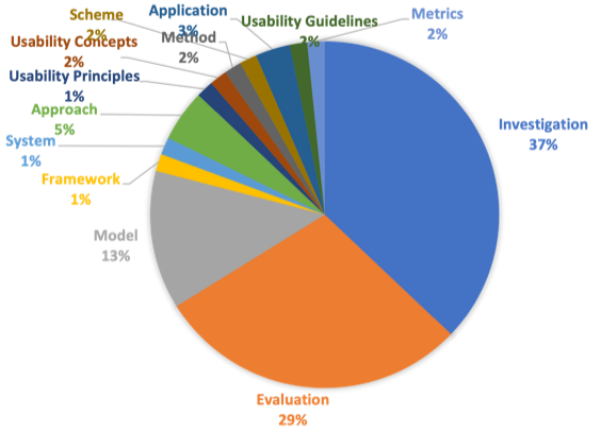

Fig. 3. Contributions by the Selected Studies.

Generally, we observed that a majority of the studies are not tailored to solution proposals, rather, they are focused on understanding software usability and ways to understand the factors that hinders users' acceptability and understandability of a software system. Despite the fact that software usability is not a recent research area, we observed that both investigative studies and evaluation studies are gaining attention from the researchers in this domain. This trend should be tailored into the proposition of new ways (in terms of framework, method, models, and so on) to help solve the usability issue in software engineering rather than just investigations and general evaluations. However, this is understandable because usability issues need to be understood using investigative and evaluation approaches.

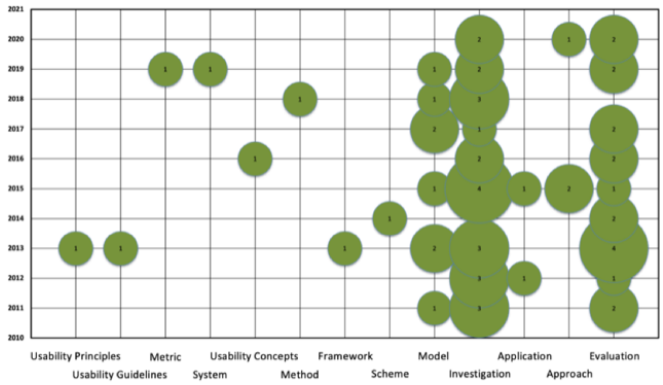

Fig. 4. Mapping of the Yearly Distribution of the Identified Contributions.

Fig. 4 shows the contributions on yearly basis. We observed that in 2018, no evaluation study was conducted. We further observed that from 2011 onward, research output is consistent and predictable. However, contributions such as framework, usability principles, usability guidelines, metric, system, usability concepts, method, scheme, and approach have received less attention in the field of study.

C. RQ3. What are the Research Types (Facet) Focused by the PS in the Domain?

From the PS, we identified four key research facets. These research facets are experience paper, evaluation research, solution proposal, and validation research. Experience papers are intended to give an explanation of how issues are handled in practice. Hence, it is generally the personal experience of the authors conducting the work. An example of the experience paper is Komiyama, et al. [42] where the authors attempted to obtain the developers point of view on the usability of intensive software system. Solution proposal seeks to provide new and novel solutions to an established problem in a research domain. The example of a solution proposal is Diaz, et al. [43]. In their work, the authors proposed a new usability metrics for e-commerce website. Evaluation research is conducted to understand how a method is implemented in practice. An example of evaluation research is the work of Al-Maani and Salameh [44]. Validation research papers present a novel proposal that is not fully implemented in practice. Examples of validation research are Störrle [45], and Christophersen and Konradt [46].

From our analysis, as presented in Table VI, we found that most of the studies conducted Evaluation research with 33% of the PS. Furthermore, experience paper constitutes 29% of the total publications, followed by solution proposal with 24%, and validation research with 13%. The analysis shows that there is an urgent need for more solution proposals and research to validate the proposed proposals. Fig. 5 depicts the map for the identified research facets in correspondence to the validation methods. We observe that more experience papers are needed in this research domain to understand users' perspective with respect to usability issues.

D. RQ4. What Validation Methods are Generally Utilized for Software Usability Evaluation?

Validation methods are key to evaluating one's work in any scientific work. In answering this RQ, we identify eight validation methods utilized by the respective PS. These methods are experiment (22 studies), theoretical validation (10), case study (5), questionnaire (3), interview (3), interview and questionnaire (2), simulation (2), and field study (1). We observed that experiment and theoretical validation are the most used approaches by the PS in this domain. In Table VII, the identified validation methods with respect to the studies that used them are highlighted.

TABLE VI. RESEARCH FACETS

Research Facet	Studies	No. of Studies	%
Evaluation Research	B3, B40, B52, B54, B50, B61, B57, B16, B44, B7, B45, B32, B55, B2, B23, B30, B31, B33, B38, B46, B49	21	33%
Experience Paper	B5, B36, B41, B15, B43, B58, B59, B4, B29, B9, B24, B11, B14, B17, B20, B34, B51, B62	18	29%
Solution Proposal	B6, B37, B48, B19, B42, B26, B39, B53, B1, B12, B13, B22, B47, B56, B60	15	24%
Validation Research	B10, B18, B21, B25, B27, B28, B8, B35	8	13%

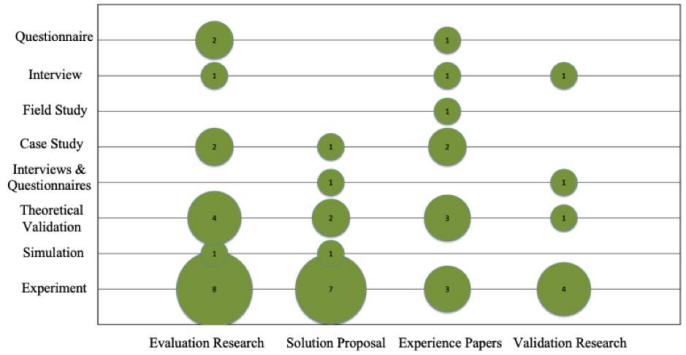

Fig. 5. Map for Research Facets against Validation Methods.

TABLE VII. VALIDATION METHODS

Research Facet	Studies	No. of Studies
Experiment	B46, B22, B60, B1, B30, B31, B21, B9, B10, B18, B29, B25, B56, B47, B2, B33, B4, B23, B49, B12, B13, B38	22
Not clearly defined	B57, B53, B26, B35, B61, B50, B51, B17, B11, B62, B34, B20, B14, B39	14
Theoretical Validation	B19, B42, B41, B36, B52, B3, B40, B27, B5, B54	10
Case Study	B59, B45, B58, B37, B7	5
Questionnaire	B15, B44, B16	3
Interview	B24, B8, B32	3
Interview and Questionnaire	B6, B28	2
Simulation	B48, B55	2
Field Study	B43	1

E. RQ5. What are the Various Application Domains for Usability?

In answering this RQ, we identified six application domains in the research area, which are web (17 studies), software development (16), mobile (8), industry and research (6), navigation (3), and robotics (1). We identified 11 studies that have no well-defined domain. Web is the most considered domain with 27% of the PS, followed by software development (25%). Table VIII presents the application domains in the research area.

TABLE VIII. APPLICATION DOMAINS

Application Domain	Studies	No. of Studies
Web	B6, B15, B8, B55, B37, B7, B30, B21, B18, B29, B56, B2, B49, B13, B38, B3, B61	17
Software Development	B43, B44, B22, B60, B10, B42, B41, B40, B27, B54, B53, B26, B51, B11, B34, B14	16
No Clear Domain Identified	B45, B58, B48, B24, B28, B12, B36, B57, B35, B50, B62	11
Mobile	B32, B46, B1, B31, B47, B33, B19, B52	8
Industry & Research	B16, B9, B4, B5, B17, B20	6
Navigation	B59, B25, B23	3
Robotics	B39	1

F. RQ6. What are the Evaluation Measures used by the PS?

With respect to evaluation metrics, we observed that most of the studies (16) used usability as a metric. This is followed by effectiveness (4 studies) and learnability (4 studies). This research observed that some studies did not clearly identified the evaluation metric that were utilized by the research. This observation is quite alarming, because almost of the primary studies have no well-defined evaluation measures in their studies. Hence, this need to be address and more work need to clarify or use an evaluation metric for their study.

V. DISCUSSION

The discussion in this section comprises of two facets. Firstly, the main findings of the study are summarized and presented clearly. It is followed by highlighting areas in usability that have not received considerable attention from the researchers.

From 2011 - 2020, the research in software usability is generally stable. Most of the PS are published in journals (39), followed by conferences (21), workshop and symposium (1 each). We found that International Journal of Human-Computer Interaction and the International Journal of Human-Computer Studies are the venues that contributed the most with 7 and 6 publications, respectively. With respect to contribution facets, multiple key contributions were identified. 37% of the PS conducted an investigative study on software usability, followed by evaluation with 29% of the PS, and model with 13%, respectively.

From our analysis, we observed that majority of the studies focused on usability testing of existing web applications (such as B6, B15, B8, B38) or usability evaluation models (such as B37, B57, B53). Other studies focused on proposing new models to help in understanding key factors that facilitate or hinder usability of software or web application. We observed that a majority of the studies are not tailored to solution proposals, rather, they are anchored in the direction of understanding software usability and ways to understand the factors that hinders users' acceptability and understandability of a software system.

From the PS, we identified four key research facets which are evaluation papers (33%), experience papers (29%), solution proposal (24%), and validation research (13%). We also identified eight validation methods utilized by the respective PS. These methods are experiment with 22 studies, followed by theoretical validation (10), case study (5), questionnaire (3), interview (3), interview and questionnaire (2), simulation (2), and field study (1). The research also identified three major evaluation metrics for evaluating proposals, which are used in this area. These metrics are usability, effectiveness and learnability. The proposals have been evaluated using a combination of metrics and sub-metrics as well. Some of the primary studies did not use any evaluation measure for validation. Hence, this needs to be addressed.

From our study, we observed that although a significant research is conducted to improve the usability of the software applications, there are still a number of research directions that needed to explore in more detail. For instance, when evaluating usability of software, majority of the works relied on closed experiments, questionnaires and surveys. It is recommended that future studies should focus more on field studies where an application usability is evaluated in real world conditions. Field studies are more beneficial as these are conducted by the end users in real operational environment resulting in identification of problems that might be overshadowed in laboratory-based tests. However, conducting the field-based study can be expensive both in terms of time and monetary value. An important aspect of software testing is the reproducibility of the errors/bugs. It is important to consider reproducibility of usability testing as a dimension in usability metrics.

Usability testing are used in a variety of domain such as aviation [47, 48], banking [49], bioinformatics [20] and medicine [28]. One area that is neglected by the researcher is the usability of software used by the scientific community. Most of the software applications developed for the use of scientific community are either command line or has poor user interfaces and faces numerous usability challenges. It will be important to consider usability testing of the scientific software and design a set of standard guidelines.

Ethics is an important aspect of software engineering process [50], however ethics in usability testing has received the least consideration by the researcher. It will be of interest to examine compliance of various ethical guidelines in usability testing. The ethical dimension of the usability testing is more important when considering the applications in domains such as health and finance.

Although considerable efforts are devoted in the literature to investigate the usability of mobile applications, we could not find any studies that focused on the cross-platform usability of

various mobile applications. Generally, mobile applications are developed for different platforms (notably android and iOS). It will be helpful to investigate how the usability of mobile applications varies across different platforms.

In the recent past machine learning applications have seen widespread use in various domains [51, 52]. However, there is limited work to evaluate the usability aspect of machine learning applications [53]. As the machine learning applications are adopted for more widespread use, it is recommended to conduct usability testing of these applications. Although, open source software has gained traction and acceptance, there is little work to assess the usability of such systems [54]. It will be important to assess the usability of open source software systems.

VI. Threat to Validity

After the rigorous analysis of the primary studies, this section discusses some identified issues that can be regarded as threat to validity. These limitations are discussed as follows.

Selection bias can be regarded as an external threat to the validity of our study. With respect to selection bias, even though the selected data sources were thoroughly searched by the author, there is a possibility that some important studies might be missed. To reduce this bias, this study employs inclusion and exclusion criteria for the selection of quality papers. The study is conducted by a single author which might result in selection bias. In order to mitigate the effect of the personal selection bias, two volunteer researchers oversaw the process and provided the feedback on the selection of final research papers for the study. Therefore, the threat level of personal selection bias is mitigated.

Conclusion validity is minimized by drawing all the relationships and conclusions from the literature based on the search query and then analyzing them using statistics such as citation count, classification into various sub-fields etc. Publication bias was mitigated by searching six important data sources. We also employed forward and backward snowballing techniques to make sure that all relevant studies related to software usability is identified, properly vetted and considered.

With respect to misclassification, some primary studies showed a limited information. Hence, some information was inferred while other where classified as “Nil” during the classification process. Hence, the inadequacy of information during classification may result in bias. In such a situation, the general methodology of a given study is carefully considered including the experimental setup to infer other classification entities. Therefore, this threat is mitigated to a certain level.

The selection of keywords can be regarded as threat of construct validity. We selected generalized keywords to identify a larger set of research papers in our domain from the various databases. Although it helped in reducing the probability of missing a relevant article, it has resulted in a large number of hits. In order to ensure that only relevant papers are selected for the study, a careful approach involving two volunteers as well as a carefully crafted approach is employed ensuring selection of relevant papers for the study.

Finally, the data was retrieved from the six selected data sources from 19 November 2020 to 25 December 2020. Therefore, if the same search query is executed at any later stage, different results might be obtained. Likewise, the citation count can be different as well as the research articles might have accrued more citations since 25th December 2020.

VII. Conclusion

This work presented a mapping study that analyzed research work from 2011 to 2020 in the domain of software usability. From an initial pool of 9,874 papers, 62 papers were carefully selected based on our inclusion/exclusion criterion. This study examined the existing contributions, research facets, evaluation measures, validation methods, application domains, demographics, publication trends, and publication forums in the research domain. With respect to contributions, investigative studies and evaluation studies are the two most common approaches. We identified four key research facets, which are experience paper, solution proposal, evaluation research, and validation research. We also identified eight validation methods utilized by the respective PS. Usability, effectiveness and learnability are found to be the common evaluation metrics. Rather alarmingly, some of the primary studies have no clear evaluation metrics defined. Hence, this need to be address and more work need to clarify or use an evaluation metric for their study.

In conclusion, the aim of this mapping study was to allow researchers and experts to have a clear understanding of the general research productivity, trends and demographics that shaped the research domain of software usability. This work will help in highlighting potential opportunities for both new and experienced researchers to conduct more works with the aim of improving the research domain.

References

[1] D. Gupta, A. K. Ahlawat, A. Sharma, and J. J. P. C. Rodrigues, "Feature selection and evaluation for software usability model using modified moth-flame optimization," *Computing*, vol. 102, pp. 1503-1520, 2020/06/01 2020. https://doi.org/10.1007/s00607-020-00809-6.

[2] K. Curcio, R. Santana, S. Reinehr, and A. Malucelli, "Usability in agile software development: A tertiary study," *Computer Standards & Interfaces*, vol. 64, pp. 61-77, 2019/05/01/ 2019. https://doi.org/10.1016/j.csi.2018.12.003.

[3] K. Sagar and A. Saha, "A systematic review of software usability studies," *International Journal of Information Technology*, 2017/12/11 2017. https://doi.org/10.1007/s41870-017-0048-1.

[4] T. Alahmadi and S. Drew, "Subjective Evaluation of Website Accessibility and Usability: A Survey for People with Sensory Disabilities," presented at the Proceedings of the 14th International Web for All Conference, Perth, Western Australia, Australia, 2017. https://doi.org/10.1145/3058555.3058579.

[5] B. Aryana and T. Clemmensen, "Mobile Usability: Experiences From Iran and Turkey," *International Journal of Human–Computer Interaction*, vol. 29, pp. 220-242, 2013/03/01 2013. 10.1080/10447318.2013.765760.

[6] A. D. Nuovo, S. Varrasi, D. Conti, J. Bamsforth, A. Lucas, A. Soranzo, *et al.*, "Usability Evaluation of a Robotic System for Cognitive Testing," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea (South), 2019, pp. 588-589. https://doi.org/10.1109/HRI.2019.8673187.

[7] I. ISO, "Information technology-software product evaluation-quality characteristics and guide lines for their use," *Iso/iec is*, vol. 9126, 1991.

[8] K. Bershinsky and R. Narciso, "An Experimental Framework for Determining the Usability of Mixed Reality Interfaces for Aerospace Applications," in *AIAA Scitech 2019 Forum*, ed. https://doi.org/10.2514/6.2019-0064.

[9] O. V. Bitkina, H. K. Kim, and J. Park, "Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges," *International Journal of Industrial Ergonomics*, vol. 76, p. 102932, 2020/03/01/ 2020. https://doi.org/10.1016/j.ergon.2020.102932.

[10] M. A. Razzak and M. N. Islam, "Exploring and Evaluating the Usability Factors for Military Application: A Road Map for HCI in Military Applications," *Human Factors and Mechanical Engineering for Defense and Safety*, vol. 4, p. 4, 2020/01/10 2020. https://doi.org/10.1007/s41314-019-0032-6.

[11] H. M. Az-zahra, N. Fauzi, and A. P. Kharisma, "Evaluating E-marketplace Mobile Application Based on People At the Center of Mobile Application Development (PACMAD) Usability Model," in *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, Lombok, Indonesia, 2019, pp. 72-77. https://doi.org/10.1109/SIET48054.2019.8986067.

[12] R. P. Bringula, "Factors Affecting Web Portal Information Services Usability: A Canonical Correlation Analysis," *International Journal of Human–Computer Interaction*, vol. 32, pp. 814-826, 2016/10/02 2016. https://doi.org/10.1080/10447318.2016.1199180.

[13] M. Bures, M. Macik, B. S. Ahmed, V. Rechtberger, and P. Slavik, "Testing the Usability and Accessibility of Smart TV Applications Using an Automated Model-Based Approach," *IEEE Transactions on Consumer Electronics*, vol. 66, pp. 134-143, 2020. https://doi.org/10.1109/TCE.2020.2986049.

[14] M. Hertzum and T. Clemmensen, "How do usability professionals construe usability?," *International Journal of Human-Computer Studies*, vol. 70, pp. 26-42, 2012/01/01/ 2012. https://doi.org/10.1016/j.ijhcs.2011.08.001.

[15] W. Isa, M. R. Suhami, N. I. Safie, and S. S. Semsudin, "Assessing the usability and accessibility of Malaysia e-government website," *American Journal of Economics and Business Administration*, vol. 3, pp. 40-46, 2011. https://doi.org/10.3844/ajebasp.2011.40.46.

[16] W. P. N. H. Pathirana and D. N. Wickramaarachchi, "Software usability improvements for Generation Z oriented software application," in *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Colombo, Sri Lanka, 2019, pp. 151-157. https://doi.org/10.23919/SCSE.2019.8842779.

[17] M. V. Waardhuizen, J. McLean-Oliver, N. Perry, and J. Munko, "Explorations on Single Usability Metrics," presented at the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk, 2019. https://doi.org/10.1145/3290607.3299062.

[18] G. J. Esmeria and R. R. Seva, "Web usability: a literature review," in *DLSU Research Congress*, Manila, Philippines, 2017.

[19] A. Linja, "Review of usability assessment tools and standards," Michigan Technological University, Applied Cognitive Science and Human Factors, Michigan Technological University2019.

[20] S. Mangul, L. S. Martin, E. Eskin, and R. Blekhman, "Improving the usability and archival stability of bioinformatics software," *Genome Biology*, vol. 20, p. 47, 2019/02/27 2019. https://doi.org/10.1186/s13059-019-1649-8.

[21] M. Sure, "Questionnaires for usability: A systematic literature review," Independent thesis Advanced level (degree of Master (Two Years)), Department of Computer and Information Science, Human-Centered systems. , Linköping University, Linköping University, 2014.

[22] J. Nielsen, *Usability engineering*, 1st edition ed. Harcourt Place, 32 Jamestown Road, London, NW1 7BY, UK: Academic Press, Inc, 1993.

[23] N. Bevan, J. Carter, and S. Harker, "ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998?," in *17th International Conference, HCI International*, Los Angeles, CA, USA, 2015, pp. 143-151. https://doi.org/10.1007/978-3-319-20901-2_13.

[24] J. M. Ferreira, S. T. Acuña, O. Dieste, S. Vegas, A. Santos, F. Rodríguez, *et al.*, "Impact of usability mechanisms: An experiment on efficiency, effectiveness and user satisfaction," *Information and Software Technology*, vol. 117, p. 106195, 2020/01/01/ 2020. https://doi.org/10.1016/j.infsof.2019.106195.

[25] R. Abiri, S. Borhani, J. Kilmarx, C. Esterwood, Y. Jiang, and X. Zhao, "A Usability Study of Low-Cost Wireless Brain-Computer Interface for Cursor Control Using Online Linear Model," *IEEE Transactions on Human-Machine Systems*, vol. 50, pp. 287-297, 2020. https://doi.org/10.1109/THMS.2020.2983848.

[26] F. P. Tulinayo, P. Ssentume, and R. Najjuma, "Digital technologies in resource constrained higher institutions of learning: a study on students' acceptance and usability," *International Journal of Educational Technology in Higher Education*, vol. 15, p. 36, 2018/09/27 2018. https://doi.org/10.1186/s41239-018-0117-y.

[27] J. M. C. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," *International Journal of Medical Informatics*, vol. 79, pp. e18-e23, 2010/04/01/ 2010. https://doi.org/10.1016/j.ijmedinf.2008.12.004.

[28] I. Maramba, A. Chatterjee, and C. Newman, "Methods of usability testing in the development of eHealth applications: A scoping review," *International Journal of Medical Informatics*, vol. 126, pp. 95-104, 2019/06/01/ 2019. https://doi.org/10.1016/j.ijmedinf.2019.03.018.

[29] P. Weichbroth, "Usability of Mobile Applications: A Systematic Literature Study," *IEEE Access*, vol. 8, pp. 55563-55577, 2020. https://doi.org/10.1109/ACCESS.2020.2981892.

[30] C. K. Coursaris and D. J. Kim, "A meta-analytical review of empirical mobile usability studies," *Journal of Usability Studies*, vol. 6, pp. 117-171, 2011.

[31] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *Journal of Interaction Science*, vol. 1, p. 1, 2013/05/07 2013. https://doi.org/10.1186/2194-0827-1-1.

[32] D. Quiñones and C. Rusu, "How to develop usability heuristics: A systematic literature review," *Computer Standards & Interfaces*, vol. 53, pp. 89-122, 2017/08/01/ 2017. https://doi.org/10.1016/j.csi.2017.03.009.

[33] I. Ahmad, G. Ahmed, S. A. A. Shah, and E. Ahmed, "A decade of big data literature: analysis of trends in light of bibliometrics," *The Journal of Supercomputing*, vol. 76, pp. 3555-3571, 2020/05/01 2020. https://doi.org/10.1007/s11227-018-2714-x.

[34] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1-18, 2015/08/01/ 2015. https://doi.org/10.1016/j.infsof.2015.03.007.

[35] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, pp. 2049-2075, 2013/12/01/ 2013. https://doi.org/10.1016/j.infsof.2013.07.010.

[36] J. Z. Maitama, N. Idris, and A. Zakari, "A Systematic Mapping Study of the Empirical Explicit Aspect Extractions in Sentiment Analysis," *IEEE Access*, vol. 8, pp. 113878-113899, 2020. https://doi.org/10.1109/ACCESS.2020.3003625.

[37] S. Ouhbi, A. Idri, J. L. Fernández-Alemán, and A. Toval, "Requirements engineering education: a systematic mapping study," *Requirements Engineering*, vol. 20, pp. 119-138, 2015/06/01 2015. https://doi.org/10.1007/s00766-013-0192-5.

[38] A. Zakari, S. P. Lee, K. A. Alam, and R. Ahmad, "Software fault localisation: a systematic mapping study," *IET Software*, vol. 13, pp. 60-74, 2019. https://doi.org/10.1049/iet-sen.2018.5137.

[39] Y. Lee and K. A. Kozar, "Understanding of website usability: Specifying and measuring constructs and their relationships," *Decision Support Systems*, vol. 52, pp. 450-463, 2012/01/01/ 2012. https://doi.org/10.1016/j.dss.2011.10.004.

[40] P. Kortum and M. Sorber, "Measuring the Usability of Mobile Applications for Phones and Tablets," *International Journal of Human–Computer Interaction*, vol. 31, pp. 518-529, 2015/08/03 2015. https://doi.org/10.1080/10447318.2015.1064658.

[41] J. Mirkovic, D. R. Kaufman, and C. M. Ruland, "Supporting Cancer Patients in Illness Management: Usability Evaluation of a Mobile App," *JMIR Mhealth Uhealth*, vol. 2, 2014. https://doi.org/10.2196/mhealth.3359.

[42] T. Komiyama, S. i. Fukuzumi, M. Azuma, H. Washizaki, and N. Tsuda, "Usability of Software–Intensive Systems from Developers' Point of View," in *22nd HCII: International Conference on Human-Computer Interaction*, Copenhagen, Denmark, 2020, pp. 450-463. https://doi.org/10.1007/978-3-030-49059-1_33.

[43] E. Diaz, S. Flores, and F. Paz, "Proposal of Usability Metrics to Evaluate E-commerce Websites," in *21st HCI International Conference on Design, User Experience, and Usability. Practice and Case Studies*, Orlando, FL, USA, 2019, pp. 85-95. https://doi.org/10.1007/978-3-030-23535-2_6.

[44] D. I. Al-Maani and H. B. Salameh, "A generic model for evaluating the usability of learning management systems," presented at the Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing, Cambridge, United Kingdom, 2017. https://doi.org/10.1145/3018896.3018921.

[45] H. Störrle, "Improving model usability and utility by layered diagrams," presented at the Proceedings of the 10th International Workshop on Modelling in Software Engineering, Gothenburg, Sweden, 2018. https://doi.org/10.1145/3193954.3193958.

[46] T. Christophersen and U. Konradt, "Reliability, validity, and sensitivity of a single-item measure of online store usability," *International Journal of Human-Computer Studies*, vol. 69, pp. 269-280, 2011/04/01/ 2011. https://doi.org/10.1016/j.ijhcs.2010.10.005.

[47] J. McSorley, J. Kleber, and B. Blickensderfer, "Usability Analysis of Aviation Weather Products for General Aviation," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, pp. 1903-1907, 2019/11/01 2019. https://doi.org/10.1177/1071181319631339.

[48] H. Xue, T. Li, X. Zhang, and R. Wang, "Integrated Usability Evaluation Method for Cockpit of Civil Aircraft," in *17th International Conference on Man–Machine–Environment System Engineering*, Jinggangshan, China, 2018, pp. 745-752. https://doi.org/10.1007/978-981-10-6232-2_89.

[49] I. Aboobucker and Y. Bao, "What obstruct customer acceptance of internet banking? Security and privacy, risk, trust and website usability and the role of moderators," *The Journal of High Technology Management Research*, vol. 29, pp. 109-123, 2018/01/01/ 2018. https://doi.org/10.1016/j.hitech.2018.04.010.

[50] I. Ozkaya, "Ethics Is a Software Design Concern," *IEEE Software*, vol. 36, pp. 4-8, 2019. https://doi.org/10.1109/MS.2019.2902592.

[51] I. Ahmad, M. A. Alqarni, A. A. Almazroi, and A. Tariq, "Experimental Evaluation of Clickbait Detection Using Machine Learning Models," *Intelligent Automation And Soft Computing*, vol. 26, pp. 1335-1344, 2020. https://doi.org/10.32604/iasc.2020.013861.

[52] I. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, "Optimizing Pretrained Convolutional Neural Networks for Tomato Leaf Disease Detection," *Complexity*, vol. 2020, p. 8812019, 2020/09/23 2020. https://doi.org/10.1155/2020/8812019.

[53] F. Bernardo, M. Zbyszyński, M. Grierson, and R. Fiebrink, "Designing and Evaluating the Usability of a Machine Learning API for Rapid Prototyping Music Technology," *Frontiers in Artificial Intelligence*, vol. 3, 2020-April-03 2020. https://doi.org/10.3389/frai.2020.00013.

[54] K. A. Dawood, K. Y. Sharif, A. A. Ghani, H. Zulzalil, A. A. Zaidan, and B. B. Zaidan, "Towards a unified criteria model for usability evaluation in the context of open source software based on a fuzzy Delphi method," *Information and Software Technology*, vol. 130, p. 106453, 2021/02/01/ 2021. https://doi.org/10.1016/j.infsof.2020.106453.

APPENDIX A. LIST OF PRIMARY STUDIES

ID	Research Article Reference
B1	Bures, M., Macik, M., Ahmed, B. S., Rechtberger, V., & Slavik, P. (2020). Testing the usability and accessibility of smart tv applications using an automated model-based approach. IEEE transactions on consumer electronics, 66(2), 134-143.
B2	Alahmadi, T., & Drew, S. (2017, April). Subjective evaluation of website accessibility and usability: A survey for people with sensory disabilities. In Proceedings of the 14th International Web for All Conference (pp. 1-4).
B3	D Al-Maani, D. I., & Bani-Salameh, H. (2017, March). A generic model for evaluating the usability of learning management systems. In Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing (pp. 27-1).
B4	Van Waardhuizen, M., McLean-Oliver, J., Perry, N., & Munko, J. (2019, May). Explorations on Single Usability Metrics. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-8).
B5	Komiyama, T., Fukuzumi, S. I., Azuma, M., Washizaki, H., & Tsuda, N. (2020, July). Usability of Software–Intensive Systems from Developers' Point of View. In International Conference on Human-Computer Interaction (pp. 450-463).
B6	Diaz, E., Flores, S., & Paz, F. (2019, July). Proposal of usability metrics to evaluate e-commerce websites. In International Conference on Human-Computer Interaction (pp. 85-95).
B7	Komarkova, J., Sedlak, P., Habrman, J., & Cermakova, I. (2017, July). Usability evaluation of web-based GIS by means of a model. In 2017 international conference on information and digital technologies (pp. 191-197).
B8	Alotaibi, K. J. (2017, May). Gathering of usability requirements by Saudi e-learning software developers. In 2017 8th International Conference on Information Technology (pp. 255-261).
B9	Okike, E. U., & Morogosi, M. (2017, July). Measuring the usability probability of learning management software using logistic regression model. In 2017 Computing Conference (pp. 1217-1223).
B10	Störrle, H. (2018, May). Improving model usability and utility by layered diagrams. In Proceedings of the 10th International Workshop on Modelling in Software Engineering (pp. 59-66).
B11	Kwon, H., & Choi, W. (2018, June). A Feld Study on Improving the API Usability of Software Platforms for Consumer Electronics Devices. In 2018 IEEE International Conference on Consumer Electronics-Asia (pp. 206-212).
B12	Geszten, D., Hámornik, B. P., & Hercegfi, K. (2018, August). Exploring awareness related usability problems of collaborative software with a team usability testing approach. In 2018 9th IEEE International Conference on Cognitive Infocommunications (pp. 45-50).
B13	McClellan, M. A., Karumur, R. P., Vogel, R. I., Petzel, S. V., Cragg, J., Chan, D., ... & Geller, M. A. (2016). Designing an educational website to improve quality of supportive oncology care for women with ovarian cancer: an expert usability review and analysis. International journal of human-computer interaction, 32(4), 297-307.
B14	Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring perceived usability: The SUS, UMUX-LITE, and AltUsability. International Journal of Human-Computer Interaction, 31(8), 496-505.

B15	Bringula, R. P. (2016). Factors affecting web portal information services usability: A canonical correlation analysis. International Journal of Human–Computer Interaction, 32(10), 814-826.
B16	Shamim, A., Balakrishnan, V., Tahir, M., & Ahsan Qureshi, M. (2016). Age and domain specific usability analysis of opinion visualisation techniques. Behaviour & Information Technology, 35(8), 680-689.
B17	Hanrath, S., & Kottman, M. (2015). Use and usability of a discovery tool in an academic library. Journal of web librarianship, 9(1), 1-21.
B18	Christophersen, T., & Konradt, U. (2011). Reliability, validity, and sensitivity of a single-item measure of online store usability. International Journal of Human-Computer Studies, 69(4), 269-280.
B19	Ko, S. M., Chang, W. S., & Ji, Y. G. (2013). Usability principles for augmented reality applications in a smartphone environment. International Journal of Human-Computer Interaction, 29(8), 501-515.
B20	Leow, M. C., Wang, L. Y. K., Lau, S. H., & Tan, C. K. (2016). Usability of rpg-based learning framework. International Journal of Human-Computer Interaction, 32(8), 643-653.
B21	Lee, Y., & Kozar, K. A. (2012). Understanding of website usability: Specifying and measuring constructs and their relationships. Decision support systems, 52(2), 450-463.
B22	Carvajal, L., Moreno, A. M., Sanchez-Segura, M. I., & Seffah, A. (2013). Usability through software design. IEEE Transactions on Software Engineering, 39(11), 1582-1596.
B23	Lavie, T., Oron-Gilad, T., & Meyer, J. (2011). Aesthetics and usability of in-vehicle navigation displays. International Journal of Human-Computer Studies, 69(1-2), 80-99.
B24	Hertzum, M., & Clemmensen, T. (2012). How do usability professionals construe usability?. International Journal of Human-Computer Studies, 70(1), 26-42.
B25	Roberts, M. J., Newton, E. J., Lagattolla, F. D., Hughes, S., & Hasler, M. C. (2013). Objective versus subjective measures of Paris Metro map usability: Investigating traditional octolinear versus all-curves schematics. International Journal of Human-Computer Studies, 71(3), 363-386.
B26	Lallemand, C. (2011, June). Toward a closer integration of usability in software development: a study of usability inputs in a model-driven engineering process. In Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems (pp. 299-302).
B27	Baum, D., Bechert, S., Eisenecker, U., Meichsner, I., & Müller, R. (2020, August). Identifying Usability Issues of Software Analytics Applications in Immersive Augmented Reality. In 2020 Working Conference on Software Visualization (pp. 100-104).
B28	Abiri, R., Borhani, S., Kilmarx, J., Esterwood, C., Jiang, Y., & Zhao, X. (2020). A usability study of low-cost wireless brain-computer interface for cursor control using online linear model. IEEE Transactions on Human-Machine Systems, 50(4), 287-297.
B29	Isa, W. A. R. W. M., Suhami, M. R., Safie, N. I., & Semsudin, S. S. (2011). Assessing the usability and accessibility of Malaysia e-government website. American Journal of Economics and Business Administration, 3(1), 40-46.
B30	Rivero, L., & Conte, T. (2013). Using an empirical study to evaluate the feasibility of a new usability inspection technique for paper based prototypes of web applications. Journal of Software Engineering Research and Development, 1(1), 1-25.
B31	Moumane, K., Idri, A., & Abran, A. (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. SpringerPlus, 5(1), 1-15.
B32	Mirkovic, J., Kaufman, D. R., & Ruland, C. M. (2014). Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR mHealth and uHealth, 2(3), e3359.
B33	Az-zahra, H. M., Fauzi, N., & Kharisma, A. P. (2019, September). Evaluating E-marketplace Mobile Application Based on People at the Center of Mobile Application Development (PACMAD) Usability Model. In 2019 International Conference on Sustainable Information Engineering and Technology (pp. 72-77).
B34	Pathirana, W. P. N. H., & Wickramaarachchi, D. N. (2019, March). Software usability improvements for Generation Z oriented software application. In 2019 International research conference on smart computing and systems engineering (pp. 151-157).
B35	Hamzah, N., Mageswaran, G., Nagappan, S. D., & Chuprat, S. (2018, October). Assessing Usability of Ubiquitous Systems Using Quality Model. In 2018 Fourth International Conference on Advances in Computing, Communication & Automation (pp. 1-4).
B36	Tamimi, H., & Bensefia, A. (2018, November). Software Usability Challenges for Native Arab Users. In 2018 3rd International Conference on System Reliability and Safety (pp. 6-12).
B37	Abuqaddom, I., Alazzam, H., Hudaib, A., & Al-Zaghoul, F. (2019, June). A measurable website usability model: Case Study University of Jordan. In 2019 10th International Conference on Information and Communication Systems (pp. 83-87).
B38	Torrente, M. C. S., Prieto, A. B. M., Gutiérrez, D. A., & De Sagastegui, M. E. A. (2013). Sirius: A heuristic-based framework for measuring web usability adapted to the type of website. Journal of Systems and Software, 86(3), 649-663.
B39	Di Nuovo, A., Varrasi, S., Conti, D., Bamsforth, J., Lucas, A., Soranzo, A., & McNamara, J. (2019, March). Usability evaluation of a robotic system for cognitive testing. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (pp. 588-589).
B40	Vilbergsdottir, S. G., Hvannberg, E. T., & Law, E. L. C. (2014). Assessing the reliability, validity and acceptance of a classification scheme of usability problems (CUP). Journal of Systems and Software, 87, 18-37.
B41	Wang, W., Cheng, J., & Guo, J. L. (2020). How Do Open Source Software Contributors Perceive and Address Usability? Valued Factors, Practices, and Challenges. IEEE Software.
B42	Polgár, P. B. (2015). Using the cognitive walkthrough method in software process improvement. E-Informatica Software Engineering Journal, 9(1).

B43	Geszten, D., Hámornik, B. P., & Hercegfi, K. (2019, October). Usability evaluation of a collaborative design software in the wild. In 2019 10th IEEE International Conference on Cognitive Infocommunications (pp. 101-106).
B44	Teruel, M. A., Navarro, E., López-Jaquero, V., Montero, F., & González, P. (2014). A CSCW requirements engineering CASE tool: development and usability evaluation. Information and Software Technology, 56(8), 922-949.
B45	Lindgaard, G. (2015). Challenges to assessing usability in the wild: a case study. International Journal of Human-Computer Interaction, 31(9), 618-631.
B46	Kortum, P., & Sorber, M. (2015). Measuring the usability of mobile applications for phones and tablets. International Journal of Human-Computer Interaction, 31(8), 518-529.
B47	Mansar, S. L., Jariwala, S., Shahzad, M., Anggraini, A., Behih, N., & AlZeyara, A. (2012). A usability testing experiment for a localized weight loss mobile application. Procedia Technology, 5, 839-848.
B48	Hurtado, N., Ruiz, M., Orta, E., & Torres, J. (2015). Using simulation to aid decision making in managing the usability evaluation process. Information and Software Technology, 57, 509-526.
B49	Hasan, L., Morris, A., & Probets, S. (2012). A comparison of usability evaluation methods for evaluating e-commerce websites. Behaviour & Information Technology, 31(7), 707-737.
B50	Hua, L., & Gong, Y. (2013, July). Usability evaluation of a voluntary patient safety reporting system: Understanding the difference between predicted and observed time values by retrospective think-aloud protocols. In International Conference on Human-Computer Interaction (pp. 94-100). Springer, Berlin, Heidelberg.
B51	Bruun, A., & Stage, J. (2012, November). Training software development practitioners in usability testing: an assessment acceptance and prioritization. In Proceedings of the 24th Australian Computer-Human Interaction Conference (pp. 52-60).
B52	Aryana, B., & Clemmensen, T. (2013). Mobile usability: experiences from Iran and Turkey. International Journal of Human-Computer Interaction, 29(4), 220-242.
B53	Moorthy, J. T. S., bin Ibrahim, S., & Mahrin, M. N. R. (2013, December). Formulation of usability risk assessment model. In 2013 IEEE Conference on Open Systems (pp. 168-173).
B54	Dubey, S. K., & Rana, A. (2011). Usability estimation of software system by using object-oriented metrics. ACM SIGSOFT Software Engineering Notes, 36(2), 1-6.
B55	Erickson, W., Trerise, S., Lee, C., VanLooy, S., Knowlton, S., & Bruyère, S. (2013). The accessibility and usability of college websites: Is your website presenting barriers to potential students?. Community College Journal of Research and Practice, 37(11), 864-876.
B56	Rodríguez, F. D., Acuña, S. T., & Juristo, N. (2015). Design and programming patterns for implementing usability functionalities in web applications. Journal of Systems and Software, 105, 107-124.
B57	Mapayi, T., Olaniyan, O., Isamotu, N., & Moses, O. (2013). Evaluating usability factors in different authentication methods using artificial neural network. African Journal of Computing & ICT, 6(1), 69-78.
B58	Wale-Kolade, A. Y. (2015). Integrating usability work into a large inter-organisational agile development project: Tactics developed by usability designers. Journal of systems and software, 100, 54-66.
B59	Brown, M., Sharples, S., & Harding, J. (2013). Introducing PEGI: A usability process for the practical evaluation of Geographic Information. International journal of human-computer studies, 71(6), 668-678.
B60	Panach, J. I., Juristo, N., Valverde, F., & Pastor, O. (2015). A framework to identify primitives that represent usability within Model-Driven Development methods. Information and Software Technology, 58, 338-354.
B61	Castilla, D., Garcia-Palacios, A., Breton-Lopez, J., Miralles, I., Baños, R. M., Etchemendy, E., ... & Botella, C. (2013). Process of design and usability evaluation of a telepsychology web and virtual reality system for the elderly: Butler. International Journal of Human-Computer Studies, 71(3), 350-362.
B62	Winter, J., & Hinley, M. (2011). Examining correlations in usability data to effectivize usability testing. E-Informatica Software Engineering Journal, 5(1).

APPENDIX B. LIST OF JOURNALS AND PROCEEDINGS

IEEE transactions on consumer electronics	Journal
Proceedings of the 14th International Web for All Conference	Conference
Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing	Conference
Conference on Human Factors in Computing Systems	Conference
International Conference on Human-Computer Interaction	Conference
International conference on information and digital technologies	Conference
International Conference on Information Technology	Conference
Computing Conference	Conference
Proceedings of the 10th International Workshop on Modelling in Software Engineering	Workshop
IEEE International Conference on Consumer Electronics-Asia	Conference
9th IEEE International Conference on Cognitive Infocommunications	Conference

International journal of human-computer interaction	Journal
Behaviour & Information Technology	Journal
Journal of web librarianship	Journal
International Journal of Human-Computer Studies	Journal
Decision support systems	Journal
IEEE Transactions on Software Engineering	Journal
Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems	Symposium
2020 Working Conference on Software Visualization	Conference
IEEE Transactions on Human-Machine Systems	Journal
American Journal of Economics and Business Administration	Journal
Journal of Software Engineering Research and Development	Journal
SpringerPlus	Journal
JMIR mHealth and uHealth	Journal
International Conference on Sustainable Information Engineering and Technology	Conference
International research conference on smart computing and systems engineering	Conference
Fourth International Conference on Advances in Computing, Communication & Automation	Conference
3rd International Conference on System Reliability and Safety	Conference
10th International Conference on Information and Communication Systems	Conference
Journal of Systems and Software	Journal
14th ACM/IEEE International Conference on Human-Robot Interaction	Conference
IEEE Software	Journal
E-Informatica Software Engineering Journal	Journal
10th IEEE International Conference on Cognitive Infocommunications	Conference
Information and Software Technology	Journal
Procedia Technology	Journal
24th Australian Computer-Human Interaction Conference	Conference
2013 IEEE Conference on Open Systems	Conference
ACM SIGSOFT Software Engineering Notes	Journal
Community College Journal of Research and Practice	Journal
African Journal of Computing & ICT	Journal