

Review

Usability Measures in Mobile-Based Augmented Reality Learning Applications: A Systematic Review

Kok Cheng Lim ^{1,2,*}, Ali Selamat ^{1,3,4,5,*} , Rose Alinda Alias ⁶ , Ondrej Krejcar ⁵  and Hamido Fujita ⁷ 

¹ School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai 81310, Johor Bahru, Johor, Malaysia

² Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, Kajang 43000, Selangor, Malaysia

³ Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra, Kuala Lumpur 54100, Malaysia

⁴ Media and Games Center of Excellence (MagicX), Universiti Teknologi Malaysia, Skudai 81310, Johor Bahru, Johor, Malaysia

⁵ Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic

⁶ Azman Hashim International Business School, Universiti Teknologi Malaysia, Skudai 81310, Johor Bahru, Johor, Malaysia

⁷ Faculty of Software and Information Science, Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate 020-0693, Japan

* Correspondence: kokcheng@uniten.edu.my (K.C.L.); aselamat@utm.my (A.S.)

Received: 25 April 2019; Accepted: 30 June 2019; Published: 5 July 2019



Abstract: The implementation of usability in mobile augmented reality (MAR) learning applications has been utilized in a myriad of standards, methodologies, and techniques. The usage and combination of techniques within research approaches are important in determining the quality of usability data collection. The purpose of this study is to identify, study, and analyze existing usability metrics, methods, techniques, and areas in MAR learning. This study adapts systematic literature review techniques by utilizing research questions and Boolean search strings to identify prospective studies from six established databases that are related to the research context area. Seventy-two articles, consisting of 45 journals, 25 conference proceedings, and two book chapters, were selected through a systematic process. All articles underwent a rigorous selection protocol to ensure content quality according to formulated research questions. Post-synthesis and analysis, the output of this article discusses significant factors in usability-based MAR learning applications. This paper presents five identified gaps in the domain of study, modes of contributions, issues within usability metrics, technique approaches, and hybrid technique combinations. This paper concludes five recommendations based on identified gaps concealing potential of usability-based MAR learning research domains, varieties of unexplored research types, validation of emerging usability metrics, potential of performance metrics, and untapped correlational areas to be discovered.

Keywords: Mobile augmented reality; usability metrics; systematic process; research domains

1. Introduction

A study done by Santos et al. in 2014 showed that within 43 studied augmented reality learning environment (ARLE) systems, usability evaluation focuses on improving ease of use, satisfaction, immersion, motivation, and performance [1]. Albert and Tullis in [2] have coined “self-reported” and “performance” as the two types of metrics in usability. Most of the surveyed research works were conducted using self-reported metrics rather than performance metrics in this field of study.

Santos et. al. (2014) reported that other common tools used among the 43 reviewed ARLE research include interviews, expert reviews, and observation, though observation remained solely the only performance metrics that are used in reported tools. On the other hand, another review paper done by [3] focused on user-based experimentation in augmented reality (AR), where emphasis was given to human perception, cognition, task performance, interaction techniques, and collaborative interaction in augmented reality (AR) generally. The work in [4] subsequently presented reviewed mobile augmented reality (MAR) works in a historical chronology fashion, highlighting the evolution of MAR research. Both systematic review works done by [1,3] and the chronological review done by [4] discussed AR in a generic form with no sub-area specification. In order to further scrutinize the studies done by the authors of [1] and [3] (the two most closely related studies), this research aims to present a more detailed systematic review focusing only on MAR learning applications to find out common usability areas of application, research types, methodologies, and techniques. There has been an increasing demand for the utilization of MAR technology in recent years. The authors in [4] highlighted broader MAR research works since 1968. The authors defined the concept of mobile AR as evolving towards the notion of a “mobile device,” aka AR on a mobile device (the miniaturization of physical devices and displays) [4]. Parallel to the definition by [4], MAR, within the context of this paper, can be defined by the implementation of augmented reality technology in the mobile environment (M), where interaction is done handheld, wireless, and in real time. The definition of MAR is supported by research works presented in [5], where MAR was described as relevant due to the increasing use of mobile devices such as smartphones and tablets providing people’s needs of communication, work, entertainment, internet access, and education. Several important works like the ones presented in [6], served as motivation towards the feasibility of integrating AR into the mobile environment through related parameters, which will eventually enhance the content of technology delivery.

2. Research Method

The work carried out in this Systematic Literature Review (SLR) has followed the guidelines presented by Barbara (2017). The research methodology flowchart is shown in Figure 1. There were 4 phases in this study. The first phase focused mainly on brainstorming and clearly defining the research questions of this research. Four research questions were formulated from this phase. Phase 2 executed the search strategy for this research. The main strategy comprised a combination of 2 consequential search techniques: Automated and manual search protocols. The first automated strategy emphasized primarily on the formulation of Boolean keyword terminologies and strings. This was then followed by searching through 5 established databases that are significant to computer science, software engineering, and information technology studies from 2009 to 2018 (recent decade). A rigorous process of paper screening based on titles, keywords, and abstracts was then performed. The second strategy was the manual searching through a protocol named the “snowballing method” introduced by [7]. Both forward and backward snowballing search protocols were extended over references of identified papers from the same time duration. Then, a similar process of paper screening based on titles, keywords, and abstracts was also performed. In Phase 3, paper selection was done through a study scrutinizing a collection of potential papers. Phase 4 then performed data synthesis, where several data visualization techniques were used to summarize the evidence of selected studies.

2.1. Research Questions

The main objectives of this SLR were to study, comprehend, and summarize the evidence of current usability metrics, methods, and techniques applied in the domain of mobile augmented reality. This study also aspired to critically identify possible research gaps and areas for future opportunities in this research domain to not only implement but to possibly expand the performance of current metrics, methods, and techniques in MAR usability studies. In order to achieve these objectives, this research formulated 4 research questions (RQ) relating to the aim of this study (Table 1).

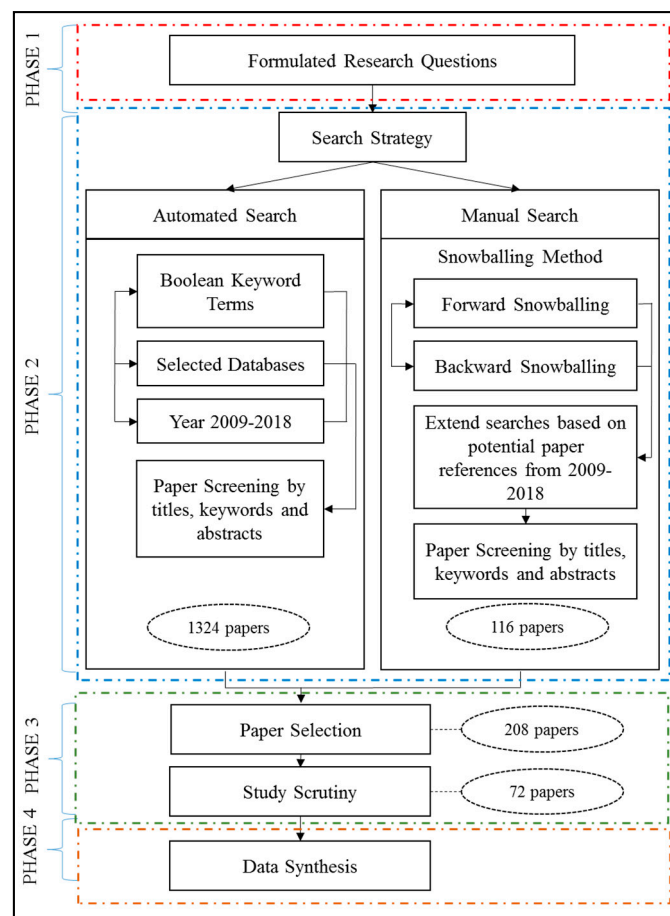


Figure 1. Research method.

Table 1. List of research questions.

Code	Research Questions
RQ1	What are the common domains, research types, and contributions for combined mobile augmented reality learning applications and usability studies?
RQ2	What are the common usability metrics used to measure usability factors of the mobile augmented reality environment?
RQ3	From the usability metrics used, what are the common methods, techniques, and instruments used in gathering usability data?
RQ4	What are the correlations in between these identified usability metrics, research types, contributions, methods, and techniques?

2.2. Search Strategies

Two main research strategies were used in this SLR (Figure 1). The strategies were executed in a sequential manner. First and foremost, an automated search was applied, and, after the identification of potential papers through the automated search, manual searches were performed only on the selected potential papers. Insights of each strategy are discussed briefly in the following section.

2.2.1. Automated Search

The automated search was based solely on a 5 point guideline suggested in [8], and it included deriving major terms, identifying alternative spelling and synonyms, keyword identification in established materials, and the utilization of Boolean operators. Since this SLR was looking forward to finding all related metrics, methods, and techniques, general terminologies relating to “usability” were used. The complete search strings are described as follows:

(Usability OR “User Experience” OR UX) AND (Mobile OR Handheld) AND (“Augmented Reality” OR AR) AND (Learn* OR Educat* OR Train*)

2.2.2. Manual Search

The manual search process for this SLR was conducted consequently after the identification of each potential paper through an automatic search. The manual search process was implemented only on the shortlisted potential papers identified through the automated search process. The manual search employed the snowballing procedure, incorporating forward and backward snowballing [9,10]. This strategy was conducted to extend the search process through references cited and papers citing the pool of potential papers. Most papers identified through these references were also initiated through a Google Scholar search, as per recommendation by [9].

2.2.3. Literature Resources

This SLR utilized searches in 6 major databases for data extraction based on the search strings formulated in Section 2.2.1. While Google Scholar was used to basically function as a triangulation mechanism in these searches, most findings in Google Scholar eventually re-directed to 5 other major databases that are significant to computer science, software engineering, and information technology studies from 2009 to 2018.

- IEEEExplore
- Web of Science
- ScienceDirect
- SpringerLink
- ACM Digital Library
- Google Scholar

2.2.4. Search Process

The search process consisted of first executing an automated search followed by the execution of a manual search process. The search was first launched in the six databases, and the papers were first collected through brief comprehension of each paper’s title, abstract, and used keywords. These potential papers were then grouped to undergo a study selection based on pre-formulated search criteria. This was then followed by an arduous analysis of the related references using the snowballing procedure, and any related papers that were marked as potential papers were included in a group of papers collected through the automated search process earlier.

2.3. Study Selection

The initial collection of the prospective papers via automated process generated a pool of 1324 papers followed by an additional 116 papers collected on a sequential manual search process. With an analysis of titles, abstracts and keywords, duplicated papers collected from different databases were deleted from the pool of papers. These papers were then classified through a pre-formulated inclusion and exclusion criteria, as presented in Table 2. After screening through this process, the papers then underwent a scrutiny of comprehensive full paper reading.

Table 2. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Include only: 1. Articles published in English language; 2. Articles with usability methods, techniques and metrics implemented; 3. Articles involving handheld mobile MAR learning applications.	Exclude all: 1. Articles published in languages other than English; 2. Articles that discuss only about application development and does not implement usability measures; 3. Articles that present study other than handheld mobile MAR learning applications.

2.3.1. Study Scrutiny

From a total of 1324 papers collected through automated search and 116 papers from manual search processes, an analysis through soft reading applied the rules of all 6 inclusion and exclusion criteria. Besides re-confirming the content of prospective papers that addressed the 4 RQs, the implementation of these 6 criteria narrowed down the number of prospective papers to a total of 208 papers, excluding papers that did not abide by one or more of the 6 criteria. All selected papers were also examined for credibility by confirming the validity of papers as originated from credible publication sources like peer-reviewed journals, conference proceedings, book chapters, and articles. From there, 208 papers underwent comprehensive readings and were coded through additional quality assessment, which was built on 5 important questions rating the suitability of these papers for this study.

Additional quality assessment questions (QAs) were implemented by answering 5 questions gauging the content of these papers (refer to Table 3). This coding method was adapted from [11], where the authors provided a credible technique for accessing the suitability of each paper for this SLR. For each QA, there were only 3 optional answers of whether the studied paper answered the QAs completely—(“Yes”), partly (“Partly”), or not at all (“No”). Each optional answer was given a coded point of either “Yes” = 1, “Partly” = 0.5, or “No” = 0. As per the practice of [7,11], QAs were done meticulously where results of the assessments were discussed by the authors prior to approval. From this process, a total of 72 papers scored 2.5 or more (50% or more), and these were selected to be included in the study of this research.

Table 3. Quality assessment questions.

QA No.	Quality Assessment Questions
1	Does the paper clearly describe the method/methods of usability used?
2	Does the paper highlight the usability evaluation process clearly?
3	Does the paper clearly present the contribution of study?
4	Does the paper clearly present the metrics used relating to types of subject study (between-subjects, within-subjects, or both)?
5	Does the paper add value to contributions towards academia, industry or community?

2.4. Data Synthesis

The main objective of data synthesis was to present and show the evidence from 72 selected studies which could assist in addressing the formulated research questions in Table 1. This process consists of data identification, synchronization, and analysis. The process will then deliver information that can clearly answer the research questions. The data collection focused on extracting areas of usability application in MAR, research types, presented contributions, usability methodologies, and techniques used. Data harvested for RQ1, RQ2, and RQ3 were organized in an articulate manner where visualization tools such as organized tables, bar charts, and pie charts were used to present the findings. Each visualization segment was paired with concluding statement descriptions to assist readers’ comprehension. RQ1, RQ2, and RQ3 also presented the respective classification of research types, presented contributions, usability methodologies, and techniques used. As for RQ4, two-dimensional and three-dimensional mappings were engineered to demonstrate correlational values between parameters in order to identify specific gaps.

3. Threats to Validity

A study by [7] highlighted four major threats to validity (TTV) in SLRs, and this research acknowledges these shortcomings by strategizing to minimize the risks of these TTVs. The four threats of validity were resolved in four separate strategies:

- Construct validity was confirmed through the implementation of an automated and manual (snowballing) search from the very beginning of data collection aimed to mitigate calculated risks. In order to further restraint this TTV, major steps of scrutiny plus additional QAs were carried out complementing existing RQs and clear selection criteria.
- Internal validity was solved by adopting a method used by [7]. In order to eliminate biases in paper selection through an exhaustive search, a combination approach of automated search and snowballing was carried out for a more inclusive selection approach. Every extracted study underwent strict selection protocols after being extracted from all major databases in similar research areas [1,7,11].
- External validity was mitigated with a generalizability of results by incorporating a 10 years' timeframe in MAR studies with a usability evaluation. The incremental collection of papers by year was parallel to the number of available papers by year, which can be an indicator that this SLR is able to maintain a generalized report aligned with the research's external validity requirements.
- Conclusion validity was managed by implementing SLR methods and techniques used in this study following the established, specific, and well-defined guidelines explored by scholars from credible publications such as [8]. It is therefore possible for each and every research chronology in this SLR to be replicated with measurable and near-identical outcomes.

4. Results and Discussion

4.1. Detailed Information of Selected Studies

In this research, a total of 72 papers from six major databases (Table 4) were selected categorized in three different clusters—45 papers (62%) came from journal publications, 25 (35%) from conference proceedings, and 2 (3%) from book chapters. From six databases, 27 papers were collected from IEEEExplore, 24 from ScienceDirect, 12 from Web of Science, 8 from Springerlink, and 1 from ACM Digital Library. While most papers were found through Google Scholar, these papers were redirected and extracted from respective source databases. Most papers were extracted from the publication year 2017 (16 paper), followed by 2018 (12 papers), 2014 (12 papers), 2015 (9 papers), 2016 (8 papers), 2013 (8 papers), 2012 (4 papers), 2011 (1 paper), 2010 (1 paper), and 2019 (1 paper). Since the paper selection process was carried out in the middle of 2018 (Figure 2), the deadline for the paper search was decided as August 2018 so that other SLR processes could be carried out as according to a planned timeline. The details of each paper are referred to in Table 5 based on publication type, publication name, year, and quartile (journal rankings when applicable).

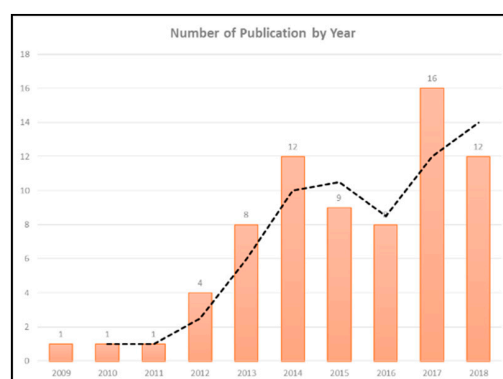


Figure 2. Publication years (duration).

Table 4. Collected articles from different online databases.

Online Databases	Articles Collected	Articles Selected after Filtering
IEEEXplore	91	27
ScienceDirect	13	24
Web of Science	53	12
SpringerLink	38	8
ACM Digital Library	13	1
Google Scholar	121	0

Table 5. List of publications.

Pub. Type	Q	Impact Factor	Year	Pub. Name	Refs.
Journal	1	1.313	2009	Journal of Computer Assisted Learning	[12]
Journal	1	1.394	2013	British Journal of Educational Technology	[13]
Journal	1	4.669	2013	Journal of Medical Internet Research	[14]
Journal	2	1.035	2013	Journal of Documentation	[15]
Journal	2	0.938	2011	Personal and Ubiquitous Computing	[16]
Journal	1	1.283	2014	IEEE Transactions on Learning Technologies	[17]
Journal	1	2.694	2014	Computers in Human Behavior	[18]
Journal	1	2.240	2014	Expert Systems with Applications	[19,20]
Journal	2	1.545	2014	IEEE Pervasive Computing	[21]
Journal	3	0.475	2014	Universal Access in the Information Society	[22]
Journal	1	1.129	2015	IEEE Transactions on Learning Technologies	[23]
Journal	1	1.330	2015	IEEE Transactions on Education	[24]
Journal	1	1.438	2015	Comunicar	[25]
Journal	1	2.880	2015	Computers in Human Behavior	[26]
Journal	1	1.719	2015	Pervasive and Mobile Computing	[27]
Journal	1	4.288	2017	IEEE Transactions on Biomedical Engineering	[28]
Journal	1	2.840	2016	IEEE Transactions on Visualization and Computer Graphics	[29]
Journal	3	NA	2016	IEEE Revista Iberoamericana de Tecnologías del Aprendizaje	[30]
Journal	1	3.977	2017	IEEE Transactions on Multimedia	[31]
Journal	1	NA	2017	Journal of Retailing and Consumer Services	[32]
Journal	1	4.538	2017	Computers and Education	[33]
Journal	1	3.129	2017	Technological Forecasting and Social Change	[34,35]
Journal	1	3.768	2017	Expert Systems with Applications	[36]
Journal	1	3.358	2017	Advanced Engineering Informatics	[37]
Journal	2	NA	2017	Entertainment Computing	[38]
Journal	2	1.541	2017	Multimedia Tools and Applications	[39]
Journal	2	1.581	2017	Microsystem Technologies	[40]
Journal	2	1.200	2017	Computers & Graphics	[41]
Journal	2	NA	2017	International Journal of Human-Computer Interaction	[42]
Journal	3	NA	2017	IEEE Revista Iberoamericana de Tecnologías del Aprendizaje	[43]
Journal	3	0.568	2015	Virtual Reality	[44]
Journal	3	NA	2017	Healthcare Technology Letters	[45]
Journal	1	4.032	2018	Automation in Construction	[46,47]
Journal	1	3.536	2018	Computers in Human Behavior	[48,49]
Journal	1	NA	2018	Journal of Retailing and Consumer Services	[50]
Journal	1	3.724	2018	Computers, Environment and Urban Systems	[51]
Journal	1	3.078	2018	IEEE Transactions on Visualization and Computer Graphics	[52]
Journal	1	2.300	2018	International Journal of Human-Computer Studies	[53]
Journal	2	1.581	2018	Microsystem Technologies	[54]
Journal	2	2.974	2018	Pervasive and Mobile Computing	[55]
Journal	NA	NA	2018	Revista Iberoamericana de Tecnologías del Aprendizaje	[56]
Proceeding	NA	NA	2010	Mobile Multimedia Processing	[57]
Proceeding	NA	NA	2012	International Symposium on Computers in Education (SIIE)	[58]
Proceeding	NA	NA	2012	Proceedings of the 2012 ACM workshop on User experience in e-learning and augmented technologies in education	[59]
Proceeding	NA	NA	2012	Procedia Computer Science	[60]
Proceeding	NA	NA	2013	Winter Simulations Conference (WSC)	[61]
Proceeding	NA	NA	2013	8th Iberian Conference on Information Systems and Technologies (CISTI)	[62]
Proceeding	NA	NA	2013	5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)	[63]
Proceeding	NA	NA	2013	Procedia Computer Science	[64]
Proceeding	NA	NA	2014	International Symposium on Computers in Education (SIIE)	[65]

Table 5. Cont.

Pub. Type	Q	Impact Factor	Year	Pub. Name	Refs.
Proceeding	NA	NA	2014	IEEE Frontiers in Education Conference (FIE) Proceedings	[66–68]
Proceeding	NA	NA	2014	Procedia Computer Science	[69]
Proceeding	NA	NA	2014	IEEE 14th International Conference on Advanced Learning Technologies	[70]
Proceeding	NA	NA	2015	IEEE 12th International Conference on e-Business Engineering	[71]
Proceeding	NA	NA	2015	International Conference on Intelligent Environments	[72]
Proceeding	NA	NA	2015	Procedia—Social and Behavioral Sciences	[73,74]
Proceeding	NA	NA	2016	13th Learning and Technology Conference (L&T)	[75]
Proceeding	NA	NA	2016	IEEE Global Engineering Education Conference (EDUCON)	[76,77]
Proceeding	NA	NA	2017	IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCM)	[78]
Proceeding	NA	NA	2017	IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)	[79]
Proceeding	NA	NA	2017	IEEE 17th International Conference on Advanced Learning Technologies (ICALT)	[80]
Proceeding	NA	NA	2017	International Conference on Orange Technologies (ICOT)	[81]
Book Chapter	NA	NA	2013	Advances in Computer Entertainment	[82]
Book Chapter	NA	NA	2016	Universal Access in Human–Computer Interaction. Interaction Techniques and Environments	[83]

In order to support the risk reduction of the aforementioned TTVs, one of the strategies was to also incorporate more papers with established quality based on measures like impact factors and journal rankings (Quartile 1 to Quartile 4). Two journal rankings tools were utilized. First, the journal details were extracted from the Thomson Reuters Master Journal List—Clarivate Analytics [84], and they were later on triangulated with SCImago Institutions rankings [85] for further information accuracy. The list of journals with impact factors is shown in Table 5 with personalized values for each journal per related year followed by the quartile rankings. A number of citations were extracted through both Google Scholar (which generally display a higher number of citations) and number of citations reported by each paper’s respective database. Studies done by Liu [12] published in the Journal of Computer Assisted Learning (Q1, Impact Factor: 1.313) appear to have the highest citation in Google Scholar (Google Scholar: 196, Publisher: 75), while studies by Olsson et al. [16] published in Personal and Ubiquitous Computing (Q2, Impact Factor: 0.938) appear to score the highest citation in the publisher’s database (SpringerLink) (Google Scholar: 181, Publisher: 87). As for proceedings, the studies done by Liu et al. in [57] appear to have had the highest number of citations in both Google Scholar and SpringerLink (Google Scholar: 77, Publisher: 19) among all collected proceedings in this research.

4.2. Domains, Research Types, and Contributions in Mobile Augmented Reality Based Usability Studies (RQ1)

In order to cater to RQ1, this section aimed to collect data and find out common domains, research types, and contributions within research works involving both MAR learning and usability studies.

4.2.1. Research Domains

Through the data synthesis carried out in this SLR, 12 categories of research domains were identified. In answering the first part of RQ1, Figure 3 shows that the education domain has the most number of MAR-based usability studies (36), followed by navigational research (15), the implementation of marketing and advertising (8), medical and health studies (3), architecture and construction (2), facility management (2), security (1), shadow emulation (1), AR gaming (1), AR visibility (1), automotive (1), and basic skills training (1). In the education domain, engineering studies (7) and architecture (7) make up most of the study population, followed by language studies (6), medical and health (2), history (2), sciences (2), and other smaller sub-domains. Other sub-domains include computer science (1) [74], primary education (1) [23], tertiary education (1) [56], mathematics (1) [62], chemistry (1) [77], animal studies (1) [81], autism (1) [21], learning in the elderly (1) [25], and safety education (1) [43] (Table 6).

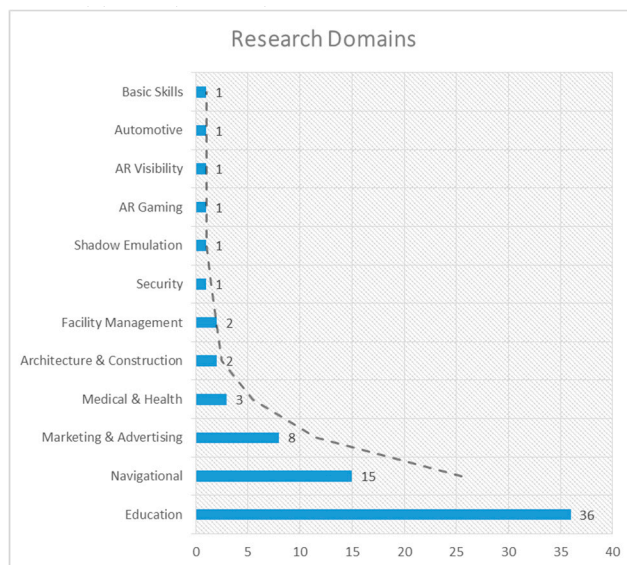


Figure 3. Research domains.

Table 6. Research domains and sub-domains.

Domain	Sub-domain	Fr.	Refs.
Education	Engineering	7	[24,26,37,61,66–68]
	Architecture	7	[18,22,58,59,63–65]
	Language	6	[12,13,57,70,75,76]
	Medical & Health	2	[14,73]
	History	2	[19,60]
	Sciences	2	[33,80]
	Others	10	[21,23,25,43,56,62,71,74,77,81]
Navigational	-	15	[15–17,27,30,39–41,48,49,53,55,69,72,79]
Marketing & Advertising	-	8	[31,32,34,35,42,50,54,83]
Medical & Health	-	3	[28,45,78]
Architecture & Construction	-	2	[46,51]
Facility Management	-	2	[20,47]
Security	-	1	[29]
Shadow Emulation	-	1	[52]
AR Gaming	-	1	[38]
AR Visibility	-	1	[44]
Automotive	-	1	[36]
Basic Skills	-	1	[82]

4.2.2. Research Types

There were five research types identified during the process of data synthesis on all 72 collected papers in answering the second part of RQ1. The categorization of research type for each paper was based on combined comprehension and understanding from experienced authors’ involved, referring to the definition of each research type discussed below:

(1) Exploratory

Exploratory research is often conducted in new areas of inquiry, where the goals of research are: (1) To scope out the magnitude or extent of a particular phenomenon, problem, or behavior; (2) to generate some initial ideas (or “hunches”) about that phenomenon; or (3) to test the feasibility of undertaking a more extensive study regarding that phenomenon. In the preliminary phases of research, when a research problem is unclear and the researcher wants to scope out the nature and extent of a certain research problem, a focus group (for individual unit of analysis) or a case study (for an organizational unit of analysis) is an ideal strategy for exploratory research [86]. According

to [87], exploratory assessments generally include thinking aloud, cognitive walkthroughs, and other techniques.

(2) Empirical

Empiricism refers to making observations to obtain knowledge. The term empirical research refers to making planned observations. By following careful plans for making observations, we engage in a systematic and thoughtful process that deserves to be called research. This process includes 7 phases namely: (1) observing; (2) selecting promising variables; (3) deciding whom to observe; (4) deciding how to observe; (5) deciding when to observe; (6) deciding how to analyze data; and (7) interpreting data to make decisions [88]. By prefixing research with empirical observations, some powerful new ideas are added. According to one definition, empirical research originates in or is based on observation or experience. Another definition holds that empirical means relying on experience or observation alone, often without due regard for system and theory [89].

(3) Comparative

The meaning of “comparative research” is restricted in accordance with what is commonly understood as comparative education in cross-national cases. It refers to “research in social units of given political level, regardless of homogeneity, similarity or difference in their cultures, although it is commonly assumed that nations always differ culturally to some degree” [90]. Traditional understandings “are compared with respect to the same concepts” [90]. In causal-comparative research, which is also called case-control research, one typically compares a group to one or more different groups, or one compares the same group at different times and does not manipulate a variable. Shadish, Cook, and Campbell argued in [91] that causal-comparative research is useful in situations in which an effect is known but the cause is unknown. For example, causal-comparative research might be used to determine what caused students to drop out of an educational program by determining how those who dropped out and those who stayed in the course differed. The key issue when running experiments is the comparison of performance between conditions: Does one condition produce better or worse performance than another? To determine “performance with condition,” human participants need to perform tasks associated with Human Computer Interaction’s (HCI) idea being investigated, and measurements of the overall performance for each condition are taken [92].

(4) Experimental

The definition of conditions, tasks, and experimental objects is the initial focus of the experiment design and must be carefully related to the research question. The experiment itself could be described simply as presenting stimuli to human participants and asking them to perform intended tasks. There are, however, many other decisions to be made about the experimental process, as well as additional supporting materials and processes to be considered [92]. In terms of experiments, scientific research may be broadly classified into two categories with slight overlap: Theoretical research and experimental research. No theory is valid until it passes one or more crucial tests of the experiment. The need for definitions in experimental research emanates from the fact that experimental researches in a given domain of nature are spread out widely over space and time [93].

(5) Quasi-Experimental

Pre-experimental and quasi-experimental research designs are often used to evaluate the effects of social programs, psychotherapy or some other form of psychosocial intervention, or the results of public policy. They are also widely used in medicine to evaluate the effects of medications. Traditionally, research designs used in outcome studies have been broadly categorized into three types. Those which involve the analysis of a single group of clients have traditionally been called pre-experimental designs. Those that involve comparing the outcomes of one group receiving a treatment that is the focus of evaluation to one or more groups of clients who receive either nothing, and alternative real treatment, or a placebo-type treatment have been called quasi-experimental designs. The third type, true experiments, are characterized by creating different groups (those receiving “real” treatment vs. those receiving nothing, alternative treatment, or placebo) by randomly assigning clients (or another unit of analysis) to those various treatment conditions. Some questions quasi-experimental can answer:

(1) What is the status of clients after they have received a given course of treatment? (2) Do clients improve after receiving a given course of treatment? (3) What is the status of clients who have received a given treatment compared to those who did not receive that treatment? (4) What is the status of clients who have received a novel treatment compared to those who received a credible placebo treatment? [94]. A quasi-experimental study might find that clients are worse following therapy, but, absent proper comparison groups, the researcher might not know that the treated clients were actually better off than if they have not received treatment [94]. In short, compared to experimental design, quasi-experimental manipulates evaluation variables but rarely has randomization in sample group (control or experimental) assignments [95]. Quasi-experimental designs are therefore more prone to bias as compared to experimental design but serve a purpose when used as a stepping stone to establish rationale of a research before subsequently leading to a conventional experimental design [95].

(6) Heuristic

According to Moustakas in [96], heuristic inquiry is a process that begins with a question or problem which the researcher seeks to illuminate or answer. The question is one that has been a personal challenge and puzzlement in the search to understand one's self and the world in which one lives. The heuristic process is autobiographic, yet, as with virtually every question that matters personally, there is also a social and perhaps universal significance. Heuristics is a way of engaging in scientific search through methods and processes aimed at discovery: A way of self-inquiry and dialogue with others aimed at finding the underlying meanings of important human experiences. The deepest currents of meaning and knowledge take place within the individual through one's senses, perceptions, beliefs, and judgements [96]. Heuristic research started out more like an informal process of assessing and meaning-making than as a research approach. Clark Moustakas, the originator of heuristic inquiry, stated that the approach came to him as he searched for a proper word to meaningfully represent certain processes he felt were foundational to explorations of everyday human experience (1990) [96]. The methodology itself was introduced in a more formalized manner to the world of research methods with the publication of Moustaka's book, in which he depicted his experience of that phenomenon as he dwelled with a decision tied to his daughter's need for heart surgery [97]. Moustakas used his personal knowledge of and relationship with loneliness as a foundation for exploring the phenomenon to others [97]. Moustaka described heuristic enquiry as a qualitative, social constructivist, and phenomenologically aligned research model [97]. In the context of social science and educational research, a heuristic enquiry has also been identified as an autobiographical approach to qualitative research. Other descriptors and characterizations of heuristic inquiry that are not highly elaborated in the professional literature include the following: Research process that studies living experience (interrelated, interconnected, and continuing experience) rather than a study of lived experienced. [97]. The word "heuristic" originates from a Greek word that means discover and explore in a wider sense [98]. Heuristics are also known as approximate techniques [98]. The main goal in a heuristic search is to construct a model that can be easily understood and that provides good solutions in a reasonable amount of computing time [98]. Such techniques consist of a combination of scientific problems such as mathematical logic, statistics, and computing, as well as human factors such as experience—and, in many cases, a good insight of the problem that needs to be addressed [98]. Heuristic design has a different perspective of research definition from the other research designs. Exploratory designs are employed when research problems are unclear in terms of scope and magnitude. Empirical research, on the other hand, uses systematically planned observation for knowledge gaining. Comparative design adopts observation of two or more competitive evaluations to derive research conclusions. Unlike the aforementioned three, experimental and quasi-experimental are preferred based on clear definition, focused conditions, tasks, and experimental objects. Heuristic, however differs from the others due to its nature of self-realization and research discoveries which take place incrementally and without systematic nor clear focused research problems.

In answering part of RQ1, a majority of 43 authors conducted the exploratory method, followed by 37 performing the comparative method, 7 performing empirical method, 6 carrying out the

experimental approach, 3 carrying out the quasi-experimental approach, and 2 performing heuristic research guidelines. Among all collected papers, 48 authors carried out only one type of research methodology, 22 carried out a combination of two research methods, and 2 carried out a combination of three research approach. Table 7 shows a complete breakdown of the research type details parameterized by a number of combinations (Comb.), research types (Type) and references (Refs). Both exploratory and comparative research methods were employed by most scholars conducting research in MAR-based usability studies. Next, this research also managed to correlate research types with the publication types including the quartile (Q) journals, non-indexed journals (NI), proceedings (P), and book chapters (BC). It can be seen in Table 7 that exploratory research was published most in Q1 journals (9), followed by Q2 journals (3), Q3 journals (1), and proceedings (12). Exploratory research appears to have been published in the highest number of high impact journals. In addition, exploratory research when adapted in combination with other research types also produced the highest number of high impact factor papers. There were 3 Q1 publications when combined with empirical research, 4 Q1 and 1 Q3 publications when combined with comparative research, and finally 3 Q1 and 1 Q2 publications when combined with experimental research. The second highest research type was comparative, when published in high impact factor publications (Q1 = 5, Q2 = 4, Q3 = 2, NI = 1, P = 8, and BC = 1). Comparative research was also the second highest when combined with other research types. In summary to all details shown in Table 7, it can be summarized that most high impact journals adapted mostly exploratory and comparative approaches.

Table 7. Combination of research types.

Comb.	Type	Refs.	Q1	Q2	Q3	NI	P	BC
1	Exploratory	[13,16,17,21–23,35,41,47,49–51,53,57,58,60–62,65,67,71,72,74,75,78]	9	3	1	-	12	-
	Empirical	-	-	-	-	-	-	-
	Comparative	[20,29,34,36,38,40,42,44,45,52,55,56,59,64,66,68,70,76,77,79,82]	5	4	2	1	8	1
	Experimental	-	-	-	-	-	-	-
	Quasi-Experimental	[30]	-	-	1	-	-	-
	Heuristic	[83]	-	-	-	-	-	1
2	Exploratory/ Empirical	[24,27,28]	3	-	-	-	-	-
	Exploratory/ Comparative	[18,31–33,43,63,80,81]	4	-	1	-	3	-
	Exploratory/ Experimental	[14,15,19,25]	3	1	-	-	-	-
	Exploratory/ Heuristic	[69]	-	-	-	-	1	-
	Empirical/ Comparative	[26,39,54]	1	2	-	-	-	-
	Comparative/ Experimental	[46]	1	-	-	-	-	-
	Comparative/ Quasi-Experimental	[12,37]	2	-	-	-	-	-
3	Exploratory/ Empirical/ Comparative	[48]	1	-	-	-	-	-
	Exploratory/ Comparative/ Experimental	[73]	-	-	-	-	1	-

Relating back to the summary of citations for each paper mentioned earlier in this paper, among all published journals, studies done by Liu [12] with the highest number of citation in Google Scholar (Google Scholar: 196, Publisher: 75) adopted a combination of comparative and experimental research approaches. On the other hand, studies done by Olsson et al. [16] with the highest number of citations

in the publisher’s database (SpringerLink) (Google Scholar: 181, Publisher: 87) appear to have adopted the exploratory research approach. Both works in [12] and [16] were published in Q1 and Q2 journals, respectively. As for proceedings and book chapters, studies done by Liu et al. in [57] with the highest number of citations in both Google Scholar and the publisher’s database (SpringerLink) (Google Scholar: 77, Publisher: 19) had also adopted the exploratory research approach.

4.2.3. Research Contributions

Referring to RQ1 from literature studies, it is important to highlight types of contribution and novelty each research paper offers. From the 72 collected papers, this research work has categorized all contribution type into five different categories defined as follows:

(1) Tool

This type of contribution focuses on producing MAR-based software tools, including systems, applications, learning packages, authoring tools, simulation tools, and prototypes, all of which can be integrated with other frameworks

(2) Method

This type of contribution concentrates on procedures and systematic processes supplementing MAR and usability research. Methods categorized here refer to learning methodologies, pedagogies, usability methodologies, and algorithms promoted with the use of MAR.

(3) Model

This type of contribution investigates the relationships, comparisons of proposed techniques, existing challenges, or classification among papers [99]. Model categorized here refers to learning models, new approaches, original concept, and innovative usability theories.

(4) Technique

This type of contribution focuses on proposing new techniques to add values to MAR research. Techniques categorized here refer to MAR technical approaches that help innovate the technology

(5) Case Study/Experience Paper

This type of contribution presents evidence on case studies and user experience involving utilization of MAR technology. Some of the contributions include exploratory findings, experimental results, user requirement studies, and comparative outcomes.

In answering the final part of RQ1, it can be seen in Table 8 that most research contributions in the area of MAR-based usability studies primarily focus on producing tools for problem solution (41), followed by formulating methods (10), designing models (9), reporting on case studies or experiences (9), and the introduction of new techniques (3).

Table 8. Types of research contribution.

Types of Contribution	Fq.	Refs.
Tool	41	[12,13,15,20–27,33,36,38,41,43,45,46,48,51,53,54,57,59,60,62–64,69–71,73–79,81–83]
Method	10	[14,18,29,30,37,44,47,56,58,65]
Model	9	[17,19,32,34,40,49,61,72,80]
Technique	3	[28,31,55]
Case Study/Experience Paper	9	[16,35,39,42,50,52,66–68]

4.3. Usability Metrics (RQ2)

RQ2 set out to find common usability metrics in measuring usability factors on the MAR learning environment. According to standards given by International Organization for Standardization (ISO) 9241-11 [100], three common metrics of measuring usability are effectiveness (accuracy and completeness in given tasks leading to objectives), efficiency (resources such as time, effort, costs, and materials to achieve goals), and satisfaction (users’ physical, cognitive, and emotional responses). However, throughout decades of usability research, these three metrics have been interpreted and varied in many different forms and terminology based on the structure recognized by ISO. In the

practice of usability metrics, they were measured through either performance metrics, self-reported metrics, or a combination of both. While the performance metric was objective (quantitative) data collection as compared to self-reported metrics—which are associated mostly as subjective (qualitative) data collections—these methods can be practiced on different sets of user groups, namely within or between-subjects. The next two sections will briefly discuss the mentioned metrics before detailing commonly used metrics in MAR-based usability studies.

4.3.1. Performance vs. Self-Reported

Tullis and Albert in [2] clearly defined performance metrics as objective methods in comparison to self-reported metrics, which are mostly subjective. Performance metrics include the usability methods that were collected mostly through observation methodologies and which do not consider the factor of participants’ opinion, while self-reported metrics only value the reliability of users’ opinions. From the 72 collected papers, there are three categories of metric segregation practiced by scholars—some only practice performance metrics, some practice only self-reported, and some combine both metrics in collecting their usability data. In answering RQ2 on most common usability metrics in MAR learning, Table 9 shows a significant majority of the authors (49) collected only self-reported data, 20 studies collected a combination of both performance and self-reported data, and only 2 authors collect pure performance data.

Table 9. Types of usability metrics category.

Types of Metrics	Fq.	Refs.
Performance	2	[28,36]
Self-reported	49	[12,13,15,16,18,19,22,25–27,31–35,37,40,41,43,49–59,61,62,64–77,80,81,83]
Combination of Both	20	[14,17,20,21,23,24,29,38,39,42,44–48,60,63,78,79,82]

4.3.2. Within-Subjects vs. Between-Subjects

According to Tullis and Albert in [2], within-subject evaluation refers to studies that performed repeated measures on experimental subjects. Commonly in usability studies, within-subject evaluation refers to having participants evaluating more than one of the tested items. The advantage of within-subject evaluation is there is no such need for a big pool of sample size—however, it risks the possibilities of the participant carryover effect and prior experience biases. Between-subjects evaluation, on the other hand, refers to comparing results for different participants, where every participant evaluates only once [2]. This evaluation type is capable of giving experimenters a clean data collection without the risks of the carryover effect and prior experience biases—however, it requires more effort and time to gather a larger sample pool. In answering another part of RQ2, Table 10 shows that a significant majority of the authors (48) performed between-subjects evaluation, 19 studies performed within-subject evaluation, and only 4 performed a combination of both.

Table 10. Sample evaluation approach.

Types of Evaluation	Fq.	Refs.
Within-subjects	19	[21,28,29,31,40,44,45,47,48,50–52,54–56,68,80–82]
Between-subjects	48	[12–19,22,24–27,32,34–39,41–43,46,49,53,57–63,65–67,69–79,83]
Combination of Both	4	[20,23,33,64]

From the collected 72 research papers, 18 categories of usability metrics were identified with multiple interchangeable terminologies in each category (Figure 4). In finalizing the answers to RQ2, it can be seen that the highest metric used is satisfaction, which goes in line with the number of majority self-reported data presented earlier. Table 11 presents the commonly used metrics together with other related terminologies within the same group. While most metrics presented below are inspired through

a derivation of three major ISO 9142-11 metrics, honorable mentions of each metrics' expression are seen important for the uniquely added values in each suggested metric.

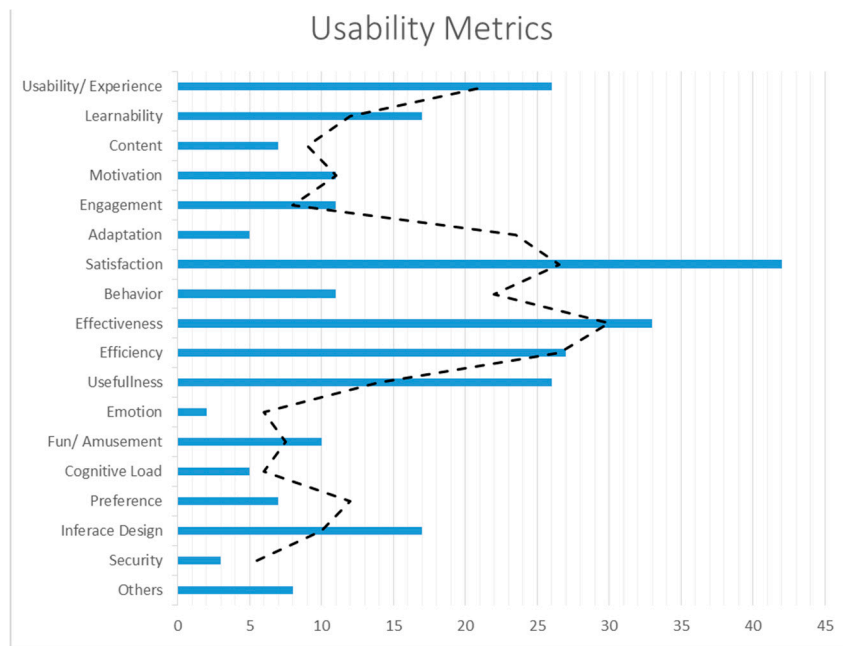


Figure 4. Groups of usability metrics.

Table 11. Usability metrics and interchangeable terminologies used by selected studies.

Metric	Interchangeable Terminologies	Refs.
Usability/ Experience	Experience	[61]
	User Experience	[14,16,19]
	Quality of experience	[39]
	Interactive experience	[42]
	Usability	[14,23,25,33,39,40,51,54,63,69,73,78,82]
	Usability ratings of severity	[69]
	User's perception	[52]
	Expectation	[16]
	Perception	[39]
	Nielsen Usability Heuristics	[83]
	Ko et al.'s MAR usability principles (five usability principles for AR)	[83]
	Usability items by (Lavie and Tractinsky) addition of (response speed and ease of control)	[42]
	Learnability	[12,23,33,38,47,48,51,81]
Learnability	Learning effectiveness	[24,63]
	Learning improvement	[73]
	Increased learning efficiency	[14]
	Education (learning)	[49]
	Learning curves	[29]
	Comprehension	[76]
	Enhancement of understanding	[73]
Content	Understandability	[44]
	Knowledge	[33]
	Perceived informativeness	[34]
	Information-feedback presentation	[40,54]
Motivation	Quality of information	[72]
	Perceived understanding	[15]
	Context awareness	[39]
Motivation	Motivation	[24,63,65,67,74]
	View angle for stimulating interest and motivating learning	[73]
Motivation	Personal innovativeness	[27]
	Behavioral intention to use	[34,37]
	Effort expectancy	[27,39]

Table 11. Cont.

Metric	Interchangeable Terminologies	Refs.
Engagement	Engagement	[33,45,49,50,53,60,61,67]
	Perceived engagement	[15]
	Emotional engagement of the different types of augmentations	[53]
	Attention (engagement)	[21]
Adaptation	Adaptation	[23]
	Comfort	[79]
	Eyestrain	[79]
	Facial expressions and body movements (Frowning, Smiling, Surprise, Concentration/Focus, Leaning close to screen)	[42]
Satisfaction	Sickness	[79]
	Satisfaction	[13,18,22,26,29,31,39–41,43–45,48–51,54,56,58–60,62,64,66,68,71,74,75,77,81]
	Perceived satisfaction	[65,70]
	Pleasure—satisfaction	[27]
	Pleasure (is happy, angry or frustrated)	[62]
	Arousal—level of satisfaction	[27]
	Factor of amusement (satisfaction)	[80]
	Satisfaction (confidence)	[38]
	User satisfaction	[35]
	Difficulty level (satisfaction)	[47]
	Overall satisfaction	[53]
	Satisfaction (exciting)	[21]
Behavior	Likeness	[29]
	Behavior	[17]
	Experimental behavior	[24]
	Behavioral Intention	[27]
	Attitude	[57]
	Perceived attitude	[32]
	Attitude towards using	[34,37]
	Appreciation	[79]
	Dominance	[27]
	Positive response	[29]
	Self-expressiveness	[39]
	Effectiveness	[12,18,20,22,29–31,33,36,37,39,44,46–48,52,57–59,64–67,71,78,81]
Effectiveness	Effectiveness (Accuracy)	[28]
	User Experience of the acceptable stability limit (effectiveness)	[55]
	Effectives—task completion	[62]
	Accuracy (performance)	[36]
	Performance expectancy	[27,39]
	Correct tasks	[48]
Efficiency	Efficiency	[18,20,22,28,29,35,36,39,44–48,58–60,64–67,81,82]
	Efficiency—understood task	[62]
	Performance (efficiency)	[38,39,43]
	productivity	[81]
Usefulness	Usefulness	[14,21,62,72,81]
	Ease of Use	[21,38,41,50,51,53,62]
	Perceived usefulness	[20,32,34,37,57,65]
	Perceived ease of use	[20,32,34,37]
Emotion	Manipulation Check (relative ease of use)	[42]
	Easiness	[57,79]
	User Friendliness	[57]
	Emotion	[14]
Fun/Amusement	Emotional Response (Arousal)	[42]
	Fun	[13,50]
	Fun (amused)	[62]
	Fun (interesting, annoying, entertaining)	[42]
	Factor of amusement (satisfaction)	[80]
	Perceived enjoyment	[32,34,37]
Fun/Amusement	Entertainment (Enjoyment)	[49]
	Negative Tone—Boring (gratifying, pleasant, confusing, and disappointing)	[42]

Table 11. Cont.

Metric	Interchangeable Terminologies	Refs.
Cognitive Load	Metacognitive Self-Regulation Skills	[24]
	Cognitive effort	[39]
	Task load	[82]
	Memories	[49]
	Labelling assist memorization	[73]
Preference	Preference	[40,44,54]
	Preferred methods of interaction	[68]
	Interest (would use again)	[62]
	Object manipulation	[73]
	Degree of interest for the content	[53]
	Aesthetics	[49,53]
	Aesthetically appreciable interface (nice)	[62]
	Attractiveness (ATT)	[14]
	Attractiveness (triggered curiosity when the instructor was presenting the Augmented Reality technology)	[62]
	Interface style	[37]
Interface Design	Presentation	[54]
	The realism of the 3-Dimensional images	[73]
	The smooth changes of images	[73]
	Realism	[79]
	Precision of 3-Dimensional images	[73]
	Quality of interface design	[72]
	Consistency	[38]
	Quality of interaction	[46]
	Simple visibility	[44]
	Universality	[81]
Security	Accessibility	[81]
	Trustfulness	[81]
	Stability	[26]
	Safety	[81]
	Escapism	[49]
Others	Facilitating conditions	[39]
	Identification (HQ-I)	[14]
	Novelty	[53]
	Pragmatic quality (PQ), hedonic	[14]
	Price value	[27]
	Social influence	[39]
	Stimulation (HQ-S), hedonic	[14]

4.4. Usability Methods, Techniques, and Instruments (RQ3)

RQ3 was formulated to find the common methods, techniques, and instruments used in gathering usability data. From the data syntheses process, there were many different usability and techniques extracted from the 72 studies collected for this research. However, rigorous analysis of the methods, techniques, and instruments used in these collected studies can be clustered into seven relative categories according to the nature of how these properties were executed. In answering part of RQ3, it is shown in Figure 5 that the questionnaire is the most used technique, registering 83 counts of usage, though some related studies use more than one type of questionnaire instruments in their studies. The next cluster is time-based tracking, which incorporates all techniques applied in the form of time collection—this registered 19 counts of usage. There were 13 occurrences of error-tracking techniques, where error counts were used to measure usability. There were 11 discussion-based techniques, which incorporate group or individual interviews. There were 10 counts of behavior tracking, 17 counts of performance-based measures, and 9 procedural or heuristics protocols in MAR-based usability studies. Table 12 shows the detail explanation of various questionnaire instruments used by selected studies parameterized by questionnaire type (Type), instruments used (Instruments), number of Likert ranges (Lik), and authors utilizing the instruments (Refs.).

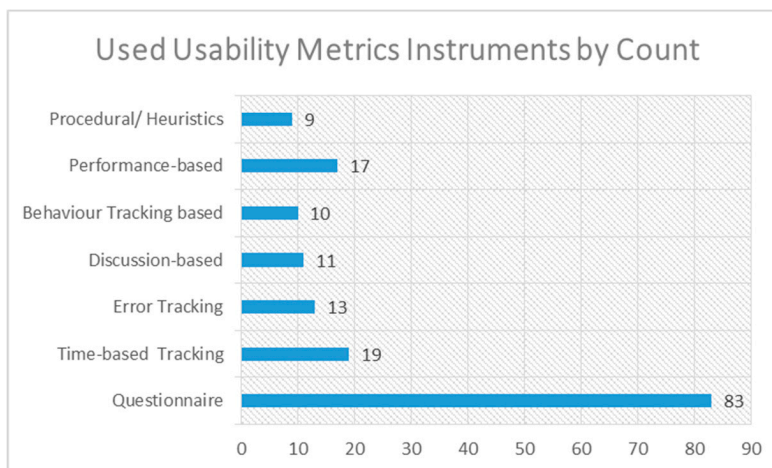


Figure 5. Used usability metrics instruments by count.

4.4.1. Open-Ended Questionnaires

According to [2], most questionnaires in usability studies include some open-ended questions in addition to the various kind of rating based questionnaires. Open-ended questionnaires are instrumental in identifying ways to improve products despite the limitations of metric calculation like close-ended questionnaires. According to [2], a common use of close-ended questionnaires in usability is to ask users for five things they liked about the product. However, summarizing the responses to this type of questions is always a challenge due to its subjectivity. Based on a definition from [101], “a questionnaire is a form designed to obtain information from respondents.” According to [102], open-ended questions are suitable for exploratory studies, supplementary to close-ended questionnaires. However, open-ended questionnaires can be demanding for respondents, require significant coding efforts, are difficult for results comparison, have a higher nonresponse rate, and require more times to answer.

4.4.2. Close-Ended Questionnaires

According to [2], even though questionnaire can be open or close-ended questions, in practical statistics for user research, most questionnaires are more typically multiple choice, with respondents selecting from a set of alternatives or points on a rating scale. According to [102], close-ended questionnaires are easier for respondents to answer since they are guided, easier to code and analysis, and appropriate when a study is certain of the possible responses. However, close-ended questionnaires can also negatively make respondents feel the absence of answers they wanted. Most closed-ended questionnaires can also be procedural, based on usability standards discussed in the next section.

4.4.3. Standardized Questionnaires

According to [101], a standardized usability questionnaire is designed for repeated use, with a specific set of questions, a specific order within a specific format, and specific rules. It is also customary for a questionnaire developer to report measurements of reliability, validity, and sensitivity of the questionnaire (psychometric qualification). There are several advantages of standardized questionnaires including objectivity, replicability, quantification, and economy. The details on advantages and ways of accessing standardized questionnaires were further referred to in [101]. According to the same source, [101], the four classic standardized usability questionnaires (used in post-study) include Questionnaire for User Interface and Satisfaction (QUIS), Software Usability Measurement Inventory (SUMI), Post-Study System Usability Questionnaire (PSSUQ), and System Usability Scale (SUS).

4.4.4. Time-Based Tracking

In this paper, all usability techniques that utilize time measures as evaluation parameters are classified in this category. According to [2], there are two common time measures that go hand in hand. One is the time of completion (also known as time-on-tasks), which refers to how quickly can users get their tasks done and how successful can these tasks be within time. According to [2], it is more reliable to register time for tasks that were done correctly, since this reflects the duration users need to perform the task given correctly.

4.4.5. Error Tracking

All usability techniques that utilize error registration as evaluation parameters are classified in this category. Sauro and Lewis in [101] defined errors as any unintended action, slip, mistake, or omission a user makes while attempting a task. According to [2], some professionals believe errors and usability issues are the same essentially. According to the same source, [2], errors are a useful way of evaluating user performance. "While being able to complete a task successfully within a reasonable amount of time is important, the number of errors made during the interaction is also very revealing." Albert and Tullis in [2] have also highlighted three general situations where error registration is useful in usability studies. According to Tullis, error is defined as entering incorrect data, making wrong menu-based choices, and failing to take a key action.

4.4.6. Discussion-Based

All usability techniques that practices user interviews and focus groups are classified in this category. Interviewing, according to [103], is a technique that favors depth over sample size, which makes it a technique that are not suitable for every problem. A focus group, on the other hand, is a moderated discussion with between four and twelve participants in a research facility, often used to explore preferences among other different solutions [103]. Both data collection methods of user interview and focus group are classified under qualitative (insights) category in a research techniques taxonomy presented in [103].

4.4.7. Expression Observation

In a taxonomy presented in [103], a research technique to gather data based on user behaviors simply means registering what people do. An example would be assessing users' behavior by reading their facial expression. In [104], it was mentioned that changes in human facial expression reflect the individual's current emotional state, which can be a means of communicating emotional information. Therefore, all usability techniques that register users' expression as a data collection approach are classified in this category.

4.4.8. Performance-Based Tracking

This category groups all usability techniques that collect data from users' performance through given experimental tasks. Tasks can be in the form of navigating a product's functionality or learning ability on the content a product is delivering. Based on insights given by [2], performance metrics rely on user behaviors and measure success based on given tasks. Performance metrics are also best used in evaluating effectiveness and efficiency [2]. Since educational assessments are oftentimes benchmarked, performance-based assessment techniques are also used due to outcome-based content standards [105].

4.4.9. Procedural

This category comprises of all reported works utilizing usability standards, procedures, heuristics, models, and protocols that have been established in the domain. According to [106], the term "standard" can be used to refer to documents approved by a recognized body or "de facto standard" that has not been approved by any recognized body but is accepted through widespread use. The importance of usability standards and procedural-based evaluation, according to [104], increase speed and cost of mobile

application development where designers do not have to reinvent wheels in development processes. Besides providing better consistency, standards also improve the quality of user experience [104].

Table 12. Usability instruments (questionnaires).

Type	Instruments	Lik	Refs.
Open-Ended	“Profile of Mood States” Questionnaire (POMS, German Variation)	-	[14]
	Questionnaires—Subject Content Performance	-	[76]
	Self-Designed Open-Ended Questionnaires	-	[12,13,17,19,20,29,38,41,45,53,56,62,72,74,75,77,79]
	Open-Ended Questionnaire for Descriptive Comments and Suggestions (34 Categories)	-	[33]
	Improved Satisfaction Questionnaire	5	[43]
	Self-Reported (Wide-Awake/Sleepy, Super Active/Passive, Enthusiastic/Apathetic, Jittery/Dull, Unaroused/Aroused) Questions based on [107–109]	5	[42]
	SFQ (Short Feedback Questionnaire) based on [110]	5	[48]
	Attrakdiff2	7	[14]
	Established Reflective Multi-Item Construct Scales from Previous Literature Questionnaire [111–114]	5	[49]
	IMMS (Keller’s Instructional Materials Motivation Survey)	5	[24]
	ISO 9241-11 Questionnaire [100]	5	[18,22,59,64,66,67]
	Nielsen’s Heuristic Evaluation & Nielsen’s Attributes Of Usability [115]	5	[67]
	Usability Satisfaction Questionnaires based on [116]	5	[67]
	The System Usability Scale (SUS) Questionnaire [117]	5	[26,40,48,58,78]
	Close-Ended	The System Usability Scale (SUS) Questionnaire [117]—Modified	5
Technology Acceptance Model (TAM) [118]		7	[20,32,37,57,67]
Technology Acceptance Model (TAM) [118]—Modified		7	[34,37]
The Motivated Strategies for Learning Questionnaires (MSLQ) [119]		5	[24]
NASA TLX Questionnaire		5	[45]
NASA TLX Questionnaire—Modified		21	[82]
Post Experiment Questionnaire’ based on Olsson [120], Designed to measure experience of MAR services		5	[46]
Post-Study System Usability Questionnaire (PSSUQ) [121]		7	[72]
Post-Study System Usability Questionnaire (PSSUQ) [121]—Modified		5	[33]
Qualitative Bipolar Laddering (BLA) Questionnaire—Test Motivation Before Use And After Use [122]		5	[65,67]
Quality of Experience (QOE) Questionnaire		5	[19]
Questionnaires based on [123]		5	[74]
Questionnaire based on [124]		5	[53]
Questionnaire Based On QUIM (Quality In Use Integrated Measurement) Factors (4) Test Data Processing To Determine Usability Percentage Value Level. [125]		5	[81]
Questionnaire Based On The Second Iteration Of The Unified Theory Of Acceptance And Use Of Technology, Which Is Commonly Referred To As UTAUT2 [126]		7	[27,39]
User Perception Questionnaire based on [127]	5	[31]	
Questionnaire for User Interface and Satisfaction—QUIS Method	5	[62]	
Self-Designed Questionnaire based on [128]	10	[35]	
Self-Designed Questionnaire (Ipsative Yes/No)	2	[41,56,79]	
Self-Designed Questionnaire by Giving 3 Separate Propositions (Not Acceptable, Acceptable, Excellent)	3	[55]	
Self-Designed Questionnaires	4	[71,77]	
Self-Designed Questionnaires	5	[16,41,42,54,61,63,68,69,73]	
Self-Designed Questionnaires	6	[23,52]	
Self-Designed Questionnaires	7	[44,79]	
Self-Designed Questionnaires	10	[52,60]	
Close-Ended Questionnaires	-	[17]	

Table 13, on the other hand, presents a myriad of usability instruments’ categories parameterized by categories by definition (Category), techniques or instruments used for measure (Techniques/ Instruments), and references of respective authors (Refs.).

Table 13. Usability instruments (pre-determined categories by definition).

Category	Technique/Instruments	Refs.	
Time-Based Tracking	Time-on-task	[20,23,29,36,39,45–48]	
	Interaction time-on-task	[63]	
	Time-on-tasks for optimal configuration	[63]	
	Task completion time	[82]	
	Number of time-on-tasks registration	[60]	
	Time-on-tasks for performance	[38]	
	Response time	[28,44]	
	Time-on-tasks across time	[48]	
	User decision time	[29]	
	Time-on-tasks for engagement	[21]	
Error Tracking	Registering the number of interaction errors	[63]	
	Number of errors for optimal configuration	[63]	
	Error rates	[44,48,82]	
	Reverse error registration	[36]	
	Error counts	[39]	
	Error registration	[28,29,46,47]	
	Absolute pose error (APE) as evaluation metrics	[28]	
	Relative pose error (RPE)	[28]	
	Interview	[12,16,23,51]	
	Interview (Interviews were transcribed and coded by two independent coders. The coders assigned a scale value (5=Strongly Agree and 1=Strongly Disagree)	[15]	
Discussion -Based	Interview for usability - only interview the teachers	[70]	
	Satisfaction interviews	[80]	
	3 rounds of mini-interviews per participant (face-to-face or video)	[50]	
	Informal interview	[48]	
	Interview (with teachers, since most students, 9 of them cannot pronounce)	[21]	
	Group discussion	[51]	
	Emotion tracking (happy, angry, unmotivated, determined) using video recording	[21]	
	Facial expression (coding through video by 2 independent coders)	[42]	
	Action and impression registration	[23]	
	Observation on student’s communication and interactivity with peers	[14]	
Behavior Observation	Observation on student’s focus on or distraction from the learning material,	[14]	
	Observation on the way students dealt with the learning object (learning material)	[14]	
	Overserving interactions	[16]	
	Engagement (switch view from mobile to non-mobile)	[45]	
	Qualitative observation by a facilitator, general tendencies in the use of a technology	[17]	
	Observing facial reaction	[80]	
	Performance-based Tracking	Pre-test and post-test on content understanding	[12,33,37,43]
		Effectiveness (task completion)	[78]
		Effectiveness (number of correct points)	[36]
		Observation-correct number of answers	[52]
User Experience of the acceptable stability limit (effectiveness)		[55]	
Effectiveness (accuracy)		[28]	
Artifact Collection (observe learning process)		[23]	
Screen recording		[23]	
Observation-video recording		[24]	
Content multiple choice		[24]	
Performance-based Tracking	Pre-test—evaluating IT and motivational profile	[65,70]	
	Observation of completion	[60]	
	Frequency of positive and negative descriptive adjectives	[34]	
	The laboratory experiments all followed the standard procedure in usability testing [129]	[34]	
	Evaluand-oriented Responsive Evaluation Model (EREM) [130]	[23]	
Procedural/ Heuristics	Cognitive walkthrough	[78]	
	Qualitative Bipolar Laddering (BLA) based on [122]	[67]	
	Heuristic (Nielsen) [131]	[69]	
	Think aloud protocol	[69]	
	Expert Reviews were used as the Nielsen heuristic evaluation (HE) method [131]	[83]	
	Ko et al.’s MAR usability principles (five usability principles for AR applications in a smart phone environment) [132]	[83]	
	Gómez et al.’s mobile-specific HE checklist [133]	[83]	

In conjunction with RQ3, this study also breaks down the more commonly used usability techniques involving observation, questionnaires, discussion-based (interviews), and procedural/heuristics (cognitive walkthroughs, think aloud, heuristics, expert reviews). Figure 6 shows the seven techniques and percentages of usage across 72 selected studies. Questionnaire (57%) is preferred most by most studies, followed by observation (23%), interviews (11%), expert review (3%), think aloud (3%), heuristic (2%), and cognitive walkthrough (1%). Among the 72 collected articles, 45 utilized only one technique at a time, 18 used a combination of two techniques, 8 utilized a combination of three techniques, and 1 study did not clearly specify specific the technique used. In Table 14, the following abbreviations apply: Questionnaire (Q), interview (Iw), observation (Obs), think aloud (TA), expert review (ER), heuristic (Hc), and cognitive walkthrough (CW). There are a maximum of three usability techniques combination at a time (Table 14). Figure 6 shows the questionnaire to be the largest pool used (57%), which is more than half, combining all the works. Figure 7 shows that although there are many techniques used, some functions are within different combination modes. Table 14 and Figure 7 show that there are 8 scholars who used a combination of three techniques in their research (Figures 8 and 9), followed by a two technique combination (18) (Figures 10 and 11). A majority of 45 scholars used only one technique at a time (Figure 12), while 1 paper did not clearly elaborate on the technique used. Figure 13 shows the frequency of all correlated techniques.

For the three technique combination, it can be comprehended that the highest used combination of techniques was that of observation, questionnaire, and interview (4), followed by the combinations of observation, questionnaire, and think aloud (1); observation, think aloud, and interview (1); observation, expert review, and interview (1); and heuristic, questionnaire, and think aloud (1). For the two technique combination, the highest combination was observation and questionnaire (14), followed by questionnaire/cognitive walkthrough (1), heuristic/expert review (1), expert analysis/interview (1), and observation/interview (1). Scholars who used only one technique in their research work fell into the following quantities: Questionnaire only (39), interviews (4), and observation (2).

Table 14. Usability technique combination.

Combination	Technique	Fq.	Refs.
Single Technique	Q	40	[13,18,19,22,26,27,31–35,37,40,41,43,49,52–59,61,62,64–68,71–77,81]
	Iw	4	[12,15,50,70]
	Obs	2	[28,36]
	Obs & Q	14	[14,20,24,29,38,39,42,44–47,60,63,79]
Combination of 2 techniques	Obs & Iw	1	[21]
	ER & Iw	1	[51]
	Hc & Er	1	[83]
	Q & CW	1	[78]
	Obs, Q & Iw	4	[17,23,48,80]
Combination of 3 techniques	Obs, TA & Iw	1	[16]
	Obs, ER & Iw	1	[25]
	Obs, Q & TA	1	[82]
	Hc, TA & Q	1	[69]
	Unclear	-	1

Note: Questionnaire (Q), interview (Iw), observation (Obs), think aloud (TA), expert review (ER), heuristic (Hc), and cognitive walkthrough (CW).

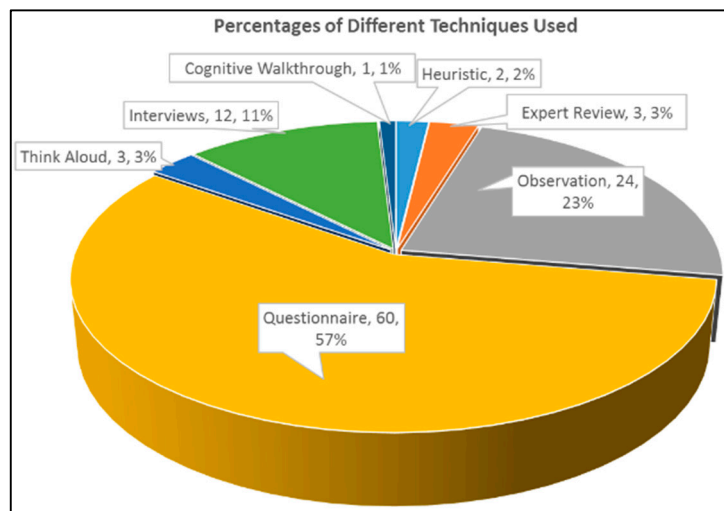


Figure 6. Percentage of usability techniques used.

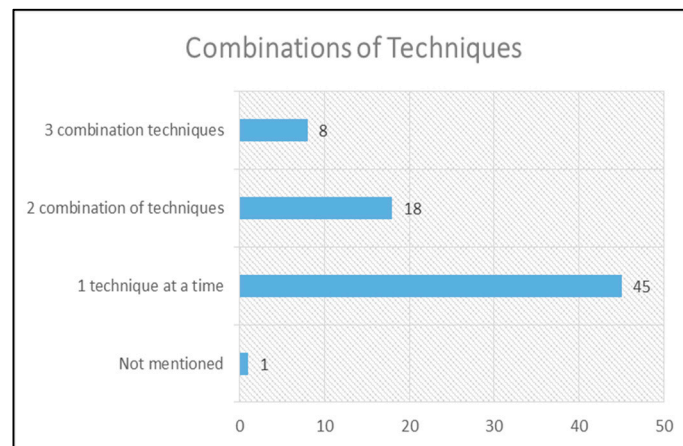


Figure 7. Research papers and usability technique combinations.

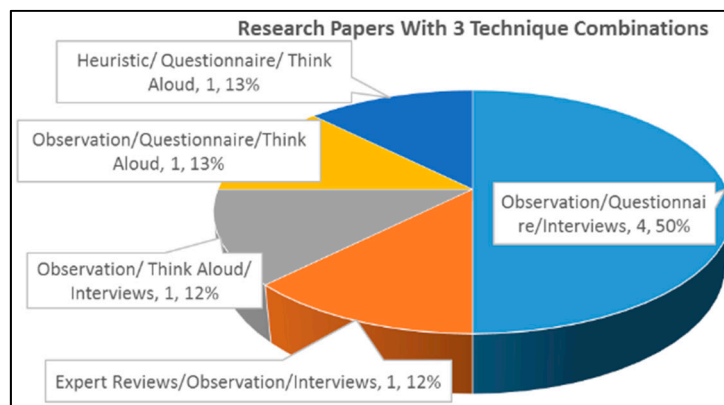


Figure 8. Papers with three technique combination.

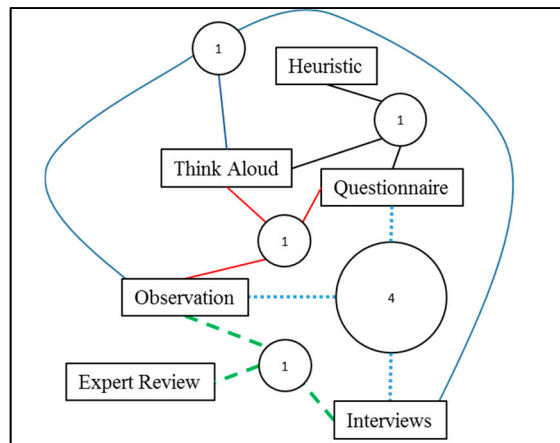


Figure 9. Works with a three technique combination.

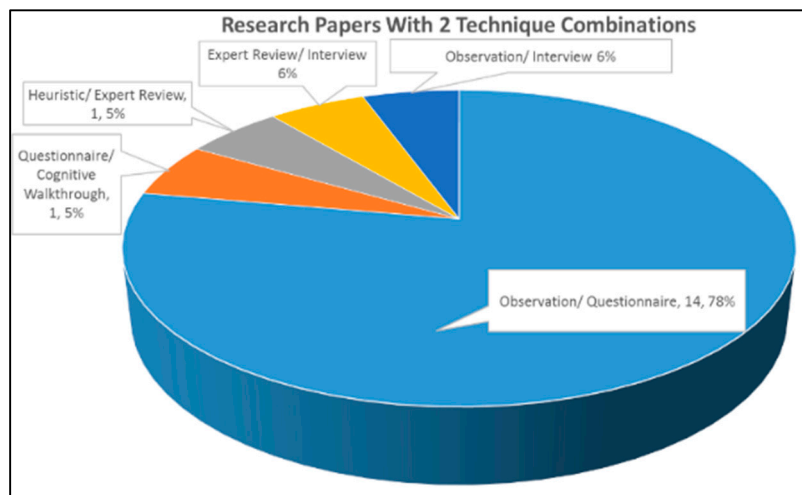


Figure 10. Papers with two technique combinations.

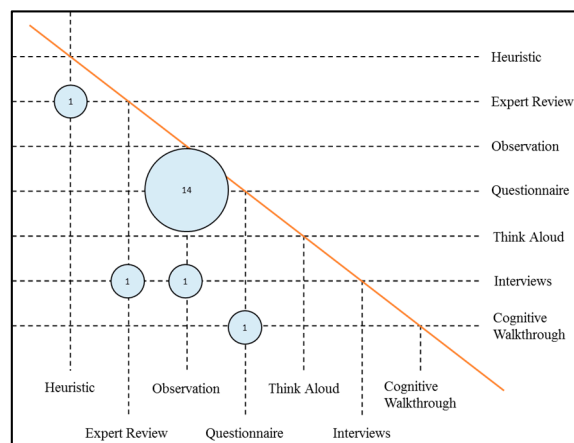


Figure 11. A number of two technique combinations.

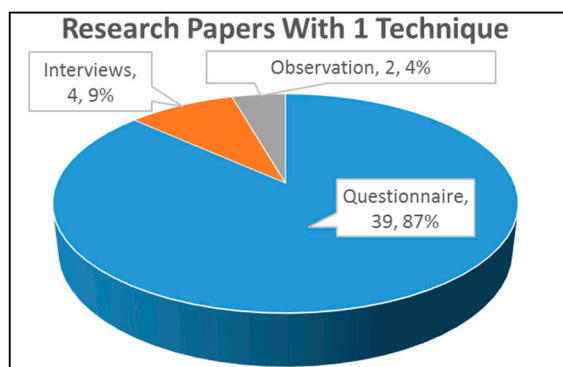


Figure 12. Papers with one main technique.

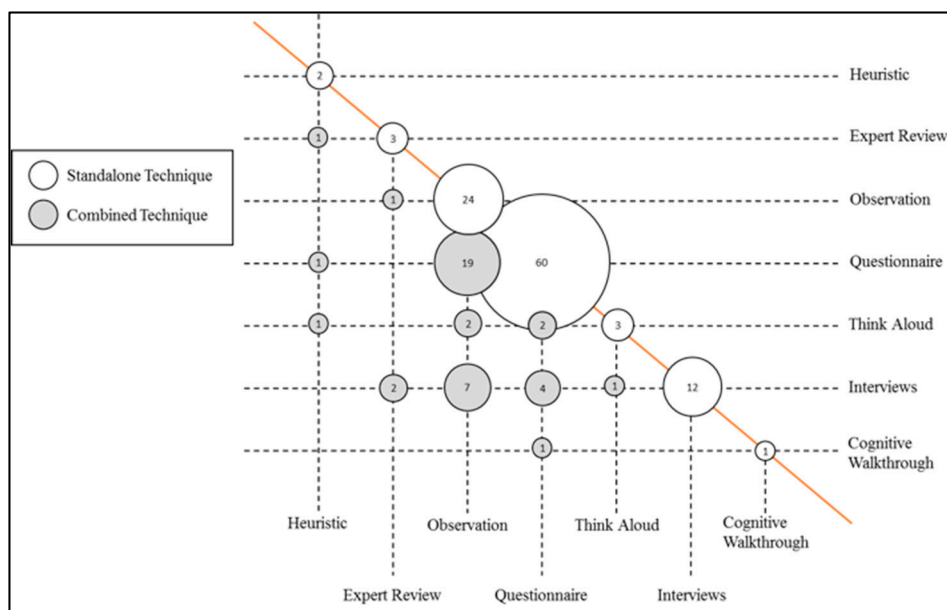


Figure 13. Frequency of correlated techniques.

4.5. Correlational Usability Mapping (RQ4)

Section 4.2, Section 4.3, Section 4.4 have, respectively, presented the common domains, research types, contributions, metrics, methods, techniques, and instruments used in MAR-based usability studies. In order to answer the fourth research question, two-dimensional and three-dimensional mappings have been employed to demonstrate co-relational factors of these collected data attributes. First, three main components—namely research types and contribution types—were mapped against the general usability metrics of performance, self-reported or combination of the two. It can be seen in Figure 14 that most studies engage in exploratory studies use self-reported metrics (32), followed by comparative studies that also utilize mostly self-reported metrics (24). The least used combination of research types and metrics are exploratory with performance (1), empirical with performance (1), and comparative with performance (1). On the other hand, most studies that contributed to the research of tools employ mostly self-reported metrics (27), followed by a hybrid of performance and self-reported metrics (13). The least used combination of contribution types and metrics are tool and performance (1), technique and performance (1), and model and hybrid metrics (1).

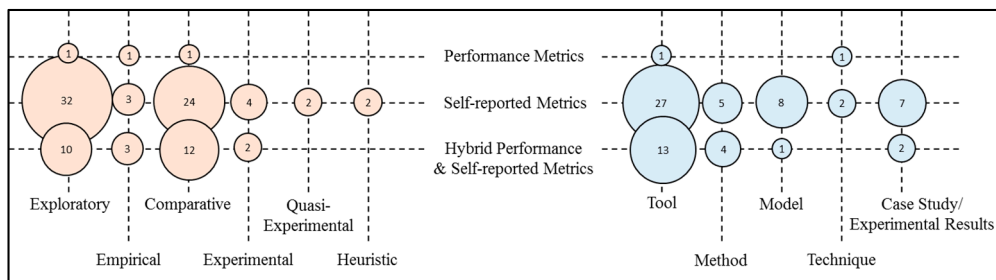


Figure 14. Two-dimensional mapping of research types, contribution types, and metrics.

The second correlation relates to a three-dimensional mapping of the mentioned metrics (performance (white), self-reported (grey) and hybrid metrics (dark grey)), research types, and contributions types. The mapping in Figure 15 shows that most studies have conducted exploratory research used self-reported metrics in the contributions of producing the MAR tool (18), followed by 12 papers that conducted comparative research using self-reported metrics in producing the MAR tool as their major contribution. Figure 16, on the other hand, presents a two-dimensional mapping of research types with seven commonly used usability techniques, followed by contribution types with the seven usability techniques. It can be seen in Figure 16 that the largest population belongs to groups of scholars who carried out exploratory research with questionnaire (36), followed by comparative research with questionnaire (34). As for techniques and tools combination, the largest pools incorporate research that combines questionnaire with tool contribution (33), followed by the combination of observation and tool contribution (14).

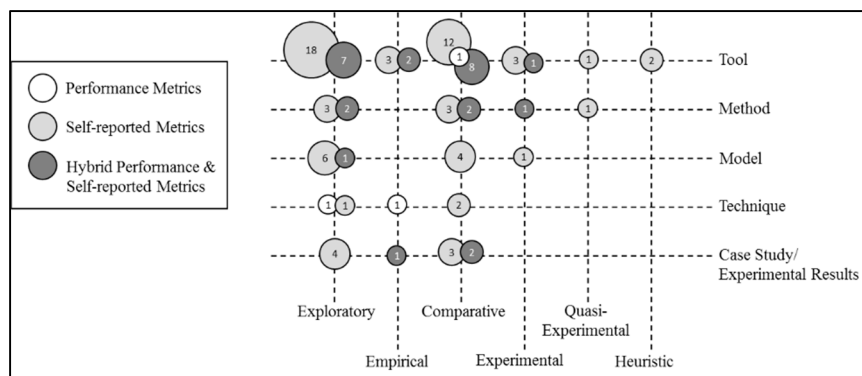


Figure 15. Three-dimensional mapping of research types, metrics, and contribution types.

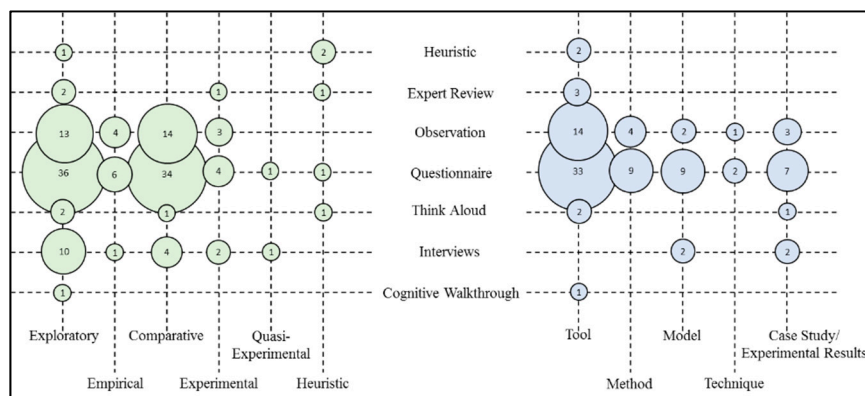


Figure 16. Two-dimensional mapping of research types, common techniques and contribution types.

Figure 17 shows a two-dimensional mapping on the correlational relationships of research types with types of evaluation and contribution type with types of evaluation. It can be interpreted from

Figure 17 that most researchers who conduct exploratory research evaluate respondents using the between-subjects technique (32), followed by comparative research, which also uses the between-subject evaluation technique (20). On the other hand, the largest pool of researchers who produce tool-related contribution also utilizes the between-subjects method (30), followed by tool-related contribution with within-subject (7), and the model-related contribution with between-subjects testing (7). It can be derived from the analysis that the largest pool of researchers uses questionnaires when measuring through between-subjects testing (41), followed by a questionnaire with within-subject testing (15). Subsequently, most scholars who use between-subjects testing performed self-reported metrics (36), followed by self-reported metrics with within-subject testing (11), and hybrid metrics with between-subjects testing (11).

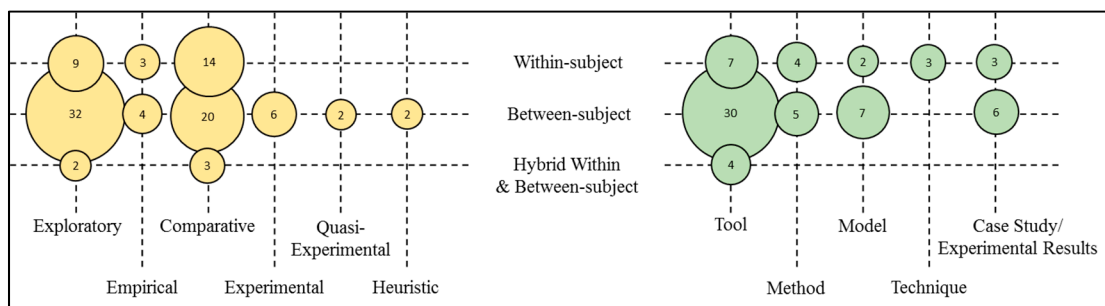


Figure 17. Two-dimensional mapping of research types with evaluation types and contribution types with evaluation types.

Referring to Figure 18, a three-dimensional mapping has been constructed to better understand the three-way correlation between common usability techniques, evaluation types and used metrics. It can be derived that the largest pool of scholars who applied questionnaire instruments in their research obviously performed self-reported metrics in a between-subjects testing fashion (30). The second largest group employed a questionnaire but also performed both metrics in a between-subjects testing setup (11). The third largest pool employed observation techniques and also carried out both metrics in a between-subjects experimental setup (10).

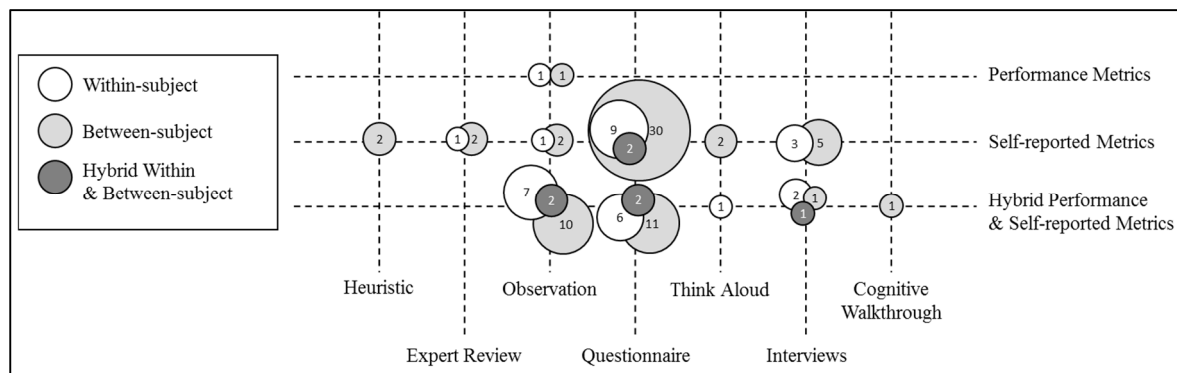


Figure 18. Three-dimensional mapping of common usability techniques with evaluation types and used metrics.

5. Research Findings on Identified Gaps

This section presents five research gaps (G1–G5) derived from results and discussion presented in Section 4 above.

5.1. Educational Domains versus Others (G1)

Figure 3 and Table 6 shows that from 12 major domain categories, the majority of the selected papers (50%) conducted usability studies on MAR in the education domain, which can be broken down into several sub-categories like engineering, architecture, and language. While performing usability studies in MAR-based educational research is promising, it can also be obsolete and seen as a complacent effort in research by focusing most studies within the education domain. While other domains, such as navigational MAR research, are catching up, the pair of MAR and usability research are still within infancy in other exploratory areas such as automotive [36], basic skills improvement [82], and AR technical research—such as works done in [44,52], gaming [38], and security [29]. As in the domains of medical health, architecture, construction, management, marketing and advertising, there have been much technical AR and applied research carried out, but has not include usability studies as one of the measured factors.

5.2. Modes of Contributions (G2)

Referring to Table 7, it can be derived that exploratory [13] and comparative [20] research types dominated among the 72 collected works. The majority of the papers produced contribution and insights relating to MAR tools. However, it is apparent that research contributions in MAR learning and usability are still lacking in the other four types of contribution. The other four types of contribution relates to research novelty on methodologies, models (especially on usability model tailored for MAR), techniques, and case studies (experience paper). Exploratory and comparative research has also been overly saturated in the area of MAR learning and usability. It is shown in the findings of this paper that there has been no research utilizing experimental and empirical approaches on their own, let alone with the combination of either exploratory or comparative studies. The utilization of quasi-experimental and heuristic methods was been minimal from the data synthesis of this research.

5.3. Standardization of Usability Metrics (G3)

While 18 categories of usability metrics performed in all 72 collected studies can be seen from Table 11 on, most used metrics are one way or another inherited from three major usability components given by ISO 9241-11 [100]. Other de facto standards like the ones recommended by prominent figures in usability studies such as Alan Dix [134] and Jakob Nielsen [115] have been instrumental in many new emerging metrics in usability domain. There are also some notable mention metrics like escapism [49], facilitating conditions [39], bundled of identification (HQ-I), pragmatic quality (PQ), stimulation (HQ-S) [14], novelty [53], price value [27], and social influence [39], which can be considered new and very much related to measures of usability from an array of different perspectives. While metrics like effectiveness, efficiency, satisfaction, and learnability are some of the most applied usability metrics across these selected papers, many other emerging metrics might or might not face validation issues since some of the mentioned metrics have yet to be accepted by the majority. Relating to G2, there are still many research loopholes in contributing to models and methodologies which can help in classifying and validating many developing usability metrics introduced in MAR. While one reference work in [83] has been identified to use procedural usability principles proposed by [132], specifically for AR in mobile environment, there is still an absence in the research of formulating standards for usability metrics in MAR, since most works adapted models and guidelines from diverse application areas.

5.4. Limited Quality versus Large Sample Convenience (G4)

One of the well-known facts in usability evaluation is the advantages and disadvantages of performance versus self-reported metrics is either risk of biases or quality of data. In usability, self-reported or subjective measures are merely means opinion-based input given by respondents expressing their experiences. They are also based heavily on subjective judgement of respondents

channeled through instruments such as questionnaires and interviews [135]. As mentioned by Olsson in [120], user experience measurements, in general, should essentially be self-reported in order to cover the subjective nature of user experience. However, according to [2,136] data collected through self-reported techniques can be subjected to social desirability bias and central tendency biases. This can lead self-reported data to be subjected to biases, inconsistencies, and validity. However, the usage of self-reported can reach larger audiences especially through scaled close-ended questionnaires. This might be the reason justifying how evidently through results of data synthesis shows 40 out of 72 scholars used only a questionnaire in tracking usability in MAR, followed by combinations of a questionnaire with other commonly used techniques. According to [101], there is an importance to highlight questionnaires reliability through several correlational approaches. While a total of 60 (57%) authors who used questionnaires as measures, only a handful (12 authors) validated the questionnaires' reliability despite risks of mentioned self-reported biases. Some of the authors who performed reliability measures includes utilization of Cronbach's Alpha [12,14,32,40,42,49,57], Cohen's Kappa [15,24], and Pearson's correlation [15,22]. Techniques such as performance metrics, which are evidently more reliable than self-reported approach were rarely used under the assumption that these processes are much more technical, time-consuming and tedious. Only a trickle pool of 24 authors (23%) executed performance metrics. A self-reported approach like questionnaire is still used primarily due to the supposition that the processes are swifter and able to reach bigger audiences compared to performance approach. Authors who performed only performance metrics like [28] reached as little as 1 sample, and [36] reached 6 samples. Authors like [32,34] only performed self-reported measures and managed to reach 978 and 318 samples, respectively. Not to mention there are also works who used a questionnaire with smaller samples, as with [62] (11 samples) and [58] (10 samples). Therefore, it is really a matter of opting for quality data within a small pool of respondents using stricter protocols or reaching a bigger audience with convenience by executing simpler procedures despite the risks of questionable and bias data collection. Hence, the identified gaps here are justifications of why self-reported metrics are still widely used despite risks of bias as compared to performance approaches.

5.5. Limitation of Hybrid Usability Methods (G5)

Figures 6–13 show the limited approaches used with usability technique combinations. As can be seen in Table 14, there are only 8 papers that promoted three technique combinations, while 18 papers proposed two technique combinations. The majority of 45 papers still preferred only one type of technique at a time. According to the ISO 9241-11 standard [100], the ideal evaluation measuring all three usability components (efficiency, effectiveness, and satisfaction) uses a combination of both performance (efficiency and effectiveness) and self-reported (satisfaction) metrics. Besides having more data angles for analysis, benefits of hybrid performance and the self-reported approach allows platforms for data triangulation, such as the counter measuring the validity of the data collected. However, perhaps due to the complexity of these procedures, less than half (26 papers) leaned on using technique combinations. The reliability of a single technique, as mentioned in G4, can be questionable, with common problems of user performances where the audience can be small and self-reported approaches that carry risks of invalid data inconsistencies. Despite the disadvantages of both approaches, some authors like [137] and [79] could still produce results with big sample audiences (50 samples or more) and report on tangible results using hybrid usability techniques.

6. Recommendations

6.1. Potential of MAR Usability in Myriad of Domains

Based on G1 and the findings presented in Section 4.2.1, MAR learning has been applied mostly in the education and navigation research areas. While there is much potential in many other areas mentioned in Section 4.2.1 and Table 6, the involvement of MAR learning in these other areas are significantly lower compared to education and navigation domain. Seeing the saturated effort of

applying combination research of MAR learning and usability in the two aforementioned areas, there are many untapped opportunities to conduct similar research penetrating into real industries aligned with requirements of Industrial Revolution 4.0.

6.2. Implementation of Research Types

Based on G2, while there is no harm in conducting more MAR-based usability studies through exploratory and comparative methodologies, and scholars might also want to look into possible study ventures using other research methodologies such empirical, experimental, quasi-experimental, and heuristic approaches since these aforementioned few have been kick-started by other researchers in the domain of MAR-based usability research. More works can also be carried out in several contribution types that are lacking in references, such as an introduction to new methodologies, models, techniques, reports of case studies, and experience papers.

6.3. Validation of New Usability Metrics in MAR

Referring to G3, there is still little-to-none research which focuses on tailoring new standards and usability metrics validation for MAR. Whilst it is crucial to produce models of usability for MAR, it is also important to validate through several recognized metrics that are relative to usability studies. Even though this SLR paper has managed to group these usability metrics in several category types (Table 11), there is still work to be done in systematically categorizing metrics within established terminologies and de facto standards through rigorous future evidence-based studies.

6.4. Utilization of Performance Metrics

It is of utmost important to highlight in G4 that performance metrics are manifestly underused despite the concrete logic of better data collection. New models and methodologies can be proposed in utilizing performance metrics that can be beneficial and at the same time eliminate commonly known limitations of objective measures. Despite having several know disadvantages in performance metrics, there are many opportunities in improving the protocols so that it can be utilized more in usability-based MAR studies. While self-reported metrics also have its set of advantages, there are also many opportunities to improve their risks on top of reliable statistically driven countermeasures.

6.5. Potential of Hybrid Techniques in MAR Usability Evaluation

The discussion of G5 has highlighted the potential of hybrid usability models that can maintain data consistency while reaching larger groups. Research and standardization work in hybrid usability approaches has opened a new gap for models and methodology introduction that can serve the objectives of improving usability in MAR. While there were several reported works in this paper on authors that utilize technique combinations, the amount of research in this area can still be improved in order to generate more result patterns achieved through hybrid usability technique in MAR.

6.6. Correlational Research

Figure 14 shows that there are still limited areas which combine hybrid usability approaches with technique contributions, the contribution of model, method, and case study/experience paper in performance metrics. Subsequently, there has also been little-to-none studies carried out using performance metrics in experimental, quasi-experimental, and heuristic research approaches. Similarly, no study had carried out hybrid metrics through quasi-experimental and heuristic approaches. Figure 17 shows that there have been no studies using an empirical approach contributing to a method or model; no studies using an experimental approach contributing to producing technique or case study/experimental results; no studies using a quasi-experimental approach contributing to model, technique, or case study/experimental results; and no studies using a heuristic approach contributing to producing method, model, technique, or case study/experimental results. Subsequently,

as also shown in Figure 17, mapping has shown that there are plenty of opportunities to investigate research type, subject study, and contribution—these include experimental with within-subject, experimental, quasi-experimental and heuristic with hybrid metrics, between-subjects studies that contribute to technique, and hybrid metrics that contribute to the method, model, technique and case study/experimental results. Figures 14–18 further elaborate on the visualization of limited research areas which could be utilized for future works in usability-based MAR studies.

7. Limitations

7.1. Quality of Work

Due to the rigorous effort of carefully comparing each paper, formulating research questions (RQ), checking through inclusion/exclusion criteria, and finally performing quality assessment questions (QAs), the confidence of the quality of work in each selected paper can be presumably high. However, there is still a risk in the definition of quality in each paper according to a different set of comprehension objectives. Despite rigorous data collection leading to a synthesis of all selected papers, we can only classify the efforts put into this article as level best and not 100% error free in assessing the quality of each selected paper. As mentioned in 3.1, automated and manual (snowballing) methods were carried out in the process in reducing inaccuracy, incompleteness, and risk of validity for data collected as much as possible.

7.2. Biases in Paper Selection

As mentioned in Section 3, both methods used in Section 7.1 were implemented to reduce biases in paper selection, but there is still no guarantee that this research has overlooked some related papers. However, there is a guarantee that all protocols had been carried out specifically to avoid any anomalies in between data collection processes.

7.3. Data Synthesis

In any review paper, external validity and conclusion validity, as discussed in Section 3, can be evident where the validity of data collected are questionable and non-general. Despite clear process implementation from the start, no processes were carried out without miscalculations, including the processes conducted in this study. Though, due to calculated risks of possible threats based on identified parameters, any errors are assumed to be minimal due to the consistent employment of SLR methods. This research followed established SLR techniques suggested by predecessor authors who had carried out similar approaches with clear evidence in minimizing the risks of TTV.

8. Conclusion

This paper aimed to study existing usability implementation in mobile augmented reality in regard to specific scope determined through four research questions. These research questions were primarily formulated to find out the existing domain of application, research types, usability metrics, methods, techniques, and approaches targeted to comprehend current issues and gaps through systematic identification. With an initial pool of 1324 papers followed by an additional 116 papers using both automated and manual searches, an arduous multi-layer process was implemented to narrow down to only 72 articles defined by pre-determined quality. Data synthesis allowed the authors in this review to understand and analyze pre-designed objectives, which eventually contributed to: (1) The classification of research demographics; (2) the categorization of usability metrics, methods, and techniques; (3) two-dimensional and three-dimensional correlational mapping between research parameters; (4) the identification of relevant research gaps; and (5) recommendations for future research in usability-based MAR derived through identified gaps and correlational mappings. The findings of this research has managed to answer the four research questions formulated earlier in this paper. RQ1 has shown evidence that the most used research domain which dominates in MAR learning is

education, followed by navigational exploration. RQ1 also highlighted the exploratory as the most adopted research type and MAR tool production as the most registered research contribution. RQ2 was answered when evidence showed self-reported metrics to be the most used usability metrics, between-subjects testing as the most preferred evaluation, and user experience to be the most measured usability parameter. In RQ3, the questionnaire was shown to be the most preferred usability techniques. Answers to RQ3 on the other hand, explained in detail the adopted combinations of usability methods, metrics, and techniques. RQ4 showed the mapping of research types, contribution types, and usability metrics from several different perspectives. Besides contributing to the detailed evidence of usability correlational variables in MAR learning, by answering all 4 RQs, this research has also managed to contribute in highlighting five research gaps addressing the varieties of related domains, a lack of contributions in several research outputs, a lack of usability standardization in MAR learning, a significant gap in usability metric utilization, and the limitation of hybrid usability methods. This paper then concluded with five recommendations founded on identified gaps in MAR learning research. The findings, synthesis, identified gaps, relational mappings, and recommendation are hoped to add value to future research works and sources that initiate more concrete studies in usability-based MAR.

Author Contributions: Conceptualization, K.C.L. and A.S.; methodology, K.C.L. and A.S.; validation, K.C.L., A.S. and O.K.; formal analysis, K.C.L. and A.S.; investigation, K.C.L. and A.S.; resources, K.C.L., A.S., R.A.A. and O.K.; data curation, K.C.L., A.S. and R.A.A.; writing—original draft preparation, K.C.L. and A.S.; writing—review and editing, K.C.L., A.S., R.A.A., O.K. and H.F.; visualization, K.C.L. and A.S.; supervision, A.S. and R.A.A.; project administration, K.C.L. and A.S.; funding acquisition, A.S. and O.K.

Funding: This research has been funded by Universiti Teknologi Malaysia (UTM) under Research University Grant Vot-20H04, Malaysia Research University Network (MRUN) Vot 4L876 and the Fundamental Research Grant Scheme (FRGS) Vot 5F073 supported under Ministry of Education Malaysia. The work is partially supported by the SPEV project, University of Hradec Kralove, FIM, Czech Republic (ID: 2102-2019). We are also grateful for the support of Ph.D. student Sebastien Mambou in consultations regarding application aspects.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Santos, M.E.C.; Chen, A.; Taketomi, T.; Yamamoto, G.; Miyazaki, J.; Kato, H. Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Trans. Learn. Technol.* **2014**, *7*, 38–56. [[CrossRef](#)]
2. Albert, W.; Tullis, T. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*; Newnes: Boston, MA, USA, 2013.
3. Edward, J.; Li, S.; Gabbard, J.L. Survey of User-Based Experimentation in augmented Reality. In Proceedings of the 1st International Conference on Virtual Reality, Las Vegas, NV, USA, 22–27 July 2005.
4. Arth, C.; Grasset, R.; Gruber, L.; Langlotz, T.; Mulloni, A.; Wagner, D. *The History of Mobile Augmented Reality*; ArXiv150501319 Cs; Cornell University: Ithaca, NY, USA, 2015.
5. Nincarean, D.; Alia, M.B.; Halim, N.D.A.; Rahman, M.H.A. Mobile augmented reality: The potential for education. *Procedia-Soc. Behav. Sci.* **2013**, *103*, 657–664. [[CrossRef](#)]
6. Liao, T. Future directions for mobile augmented reality research: Understanding relationships between augmented reality users, nonusers, content, devices, and industry. *Mob. Media Commun.* **2019**, *7*, 131–149. [[CrossRef](#)]
7. Zhou, X.; Jin, Y.; Zhang, H.; Li, S.; Huang, X. A Map of Threats to Validity of Systematic Literature Reviews in Software Engineering. In Proceedings of the 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), Hamilton, New Zealand, 6–9 December 2016; pp. 153–160.
8. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [[CrossRef](#)]
9. Wohlin, C. 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, New York, NY, USA, 13–14 May 2014; p. 38.
10. Zhang, H.; Babar, M.A.; Tell, P. Identifying relevant studies in software engineering. *Inf. Softw. Technol.* **2011**, *53*, 625–637. [[CrossRef](#)]

11. Achimugu, P.; Selamat, A.; Ibrahim, R.; Mahrin, M.N. A systematic literature review of software requirements prioritization research. *Inf. Softw. Technol.* **2014**, *56*, 568–585. [[CrossRef](#)]
12. Liu, T.-Y. A context-aware ubiquitous learning environment for language listening and speaking. *J. Comput. Assist. Learn.* **2009**, *25*, 515–527. [[CrossRef](#)]
13. Liu, P.-H.E.; Tsai, M.-K. Using augmented-reality-based mobile learning material in EFL English composition: An exploratory case study. *Br. J. Educ. Technol.* **2013**, *44*, E1–E4. [[CrossRef](#)]
14. Albrecht, U.-V.; Folta-Schoofs, K.; Behrends, M.; von Jan, U. Effects of Mobile Augmented Reality Learning Compared to Textbook Learning on Medical Students: Randomized Controlled Pilot Study. *J. Med. Internet Res.* **2013**, *15*, e182. [[CrossRef](#)]
15. Cociolo, A.; Rabina, D. Does place affect user engagement and understanding? Mobile learner perceptions on the streets of New York. *J. Doc.* **2013**, *69*, 98–120. [[CrossRef](#)]
16. Olsson, T.; Lagerstam, E.; Kärkkäinen, T.; Väänänen-Vainio-Mattila, K. Expected user experience of mobile augmented reality services: A user study in the context of shopping centres. *Pers. Ubiquitous Comput.* **2011**, *17*, 287–304. [[CrossRef](#)]
17. Pérez-Sanagustín, M.; Hernández-Leo, D.; Santos, P.; Kloos, C.D.; Blat, J. Augmenting Reality and Formality of Informal and Non-Formal Settings to Enhance Blended Learning. *IEEE Trans. Learn. Technol.* **2014**, *7*, 118–131. [[CrossRef](#)]
18. Fonseca, D.; Martí, N.; Redondo, E.; Navarro, I.; Sánchez, A. Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models. *Comput. Hum. Behav.* **2014**, *31*, 434–445. [[CrossRef](#)]
19. Blanco-Fernández, Y.; López-Nores, M.; Pazos-Arias, J.J.; Gil-Solla, A.; Ramos-Cabrera, M.; García-Duque, J. REENACT: A step forward in immersive learning about Human History by augmented reality, role playing and social networking. *Expert Syst. Appl.* **2014**, *41*, 4811–4828. [[CrossRef](#)]
20. Shatte, A.; Holdsworth, J.; Lee, I. Mobile augmented reality based context-aware library management system. *Expert Syst. Appl.* **2014**, *41*, 2174–2185. [[CrossRef](#)]
21. Escobedo, L.; Tentori, M.; Quintana, E.; Favela, J.; Garcia-Rosas, D. Using Augmented Reality to Help Children with Autism Stay Focused. *IEEE Pervasive Comput.* **2014**, *13*, 38–46. [[CrossRef](#)]
22. Riera, A.S.; Redondo, E.; Fonseca, D. Geo-located teaching using handheld augmented reality: Good practices to improve the motivation and qualifications of architecture students. *Univers. Access Inf. Soc.* **2014**, *14*, 363–374. [[CrossRef](#)]
23. Muñoz-Cristóbal, J.A.; Jorrín-Abellán, I.M.; Asensio-Pérez, J.I.; Martínez-Monés, A.; Prieto, L.P.; Dimitriadis, Y. Supporting Teacher Orchestration in Ubiquitous Learning Environments: A Study in Primary Education. *IEEE Trans. Learn. Technol.* **2015**, *8*, 83–97. [[CrossRef](#)]
24. Ibáñez, M.B.; Di-Serio, Á.; Villarán-Molina, D.; Delgado-Kloos, C. Augmented Reality-Based Simulators as Discovery Learning Tools: An Empirical Study. *IEEE Trans. Educ.* **2015**, *58*, 208–213. [[CrossRef](#)]
25. Saracchini, R.; Catalina, C.; Bordoni, L. A Mobile Augmented Reality Assistive Technology for the Elderly. *Comunicar* **2015**, *23*, 65–73. [[CrossRef](#)]
26. Martín-Gutiérrez, J.; Fabiani, P.; Benesova, W.; Meneses, M.D.; Mora, C.E. Augmented reality to promote collaborative and autonomous learning in higher education. *Comput. Hum. Behav.* **2015**, *51*, 752–761. [[CrossRef](#)]
27. Kourouthanassis, P.; Boletsis, C.; Bardaki, C.; Chasanidou, D. Tourists responses to mobile augmented reality travel guides: The role of emotions on adoption behaviour. *Pervasive Mob. Comput.* **2015**, *18*, 71–87. [[CrossRef](#)]
28. Rodas, N.L.; Barrera, F.; Padoy, N. See It with Your Own Eyes: Markerless Mobile Augmented Reality for Radiation Awareness in the Hybrid Room. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 429–440.
29. Hartl, A.D.; Arth, C.; Grubert, J.; Schmalstieg, D. Efficient Verification of Holograms Using Mobile Augmented Reality. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 1843–1851. [[CrossRef](#)] [[PubMed](#)]
30. Nagata, J.J.; Giner, J.R.G.-B.; Abad, F.M. Virtual Heritage of the Territory: Design and Implementation of Educational Resources in Augmented Reality and Mobile Pedestrian Navigation. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2016**, *11*, 41–46. [[CrossRef](#)]
31. Sekhavat, Y.A. Privacy Preserving Cloth Try-On Using Mobile Augmented Reality. *IEEE Trans. Multimed.* **2017**, *19*, 1041–1049. [[CrossRef](#)]

32. Pantano, E.; Rese, A.; Baier, D. Enhancing the online decision-making process by using augmented reality: A two country comparison of youth markets. *J. Retail. Consum. Serv.* **2017**, *38*, 81–95. [[CrossRef](#)]
33. Frank, J.A.; Kapila, V. Mixed-reality learning environments: Integrating mobile interfaces with laboratory test-beds. *Comput. Educ.* **2017**, *110*, 88–104. [[CrossRef](#)]
34. Rese, A.; Baier, D.; Geyer-Schulz, A.; Schreiber, S. How augmented reality apps are accepted by consumers: A comparative analysis using scales and opinions. *Technol. Forecast. Soc. Chang.* **2017**, *124*, 306–319. [[CrossRef](#)]
35. Dacko, S.G. Enabling smart retail settings via mobile augmented reality shopping apps. *Technol. Forecast. Soc. Chang.* **2017**, *124*, 243–256. [[CrossRef](#)]
36. Lima, J.P.; Roberto, R.; Simões, F.; Almeida, M.; Figueiredo, L.; Teixeira, J.M.; Teichrieb, V. Markerless tracking system for augmented reality in the automotive industry. *Expert Syst. Appl.* **2017**, *82*, 100–114. [[CrossRef](#)]
37. Turkan, Y.; Radkowski, R.; Karabulut-Ilgu, A.; Behzadan, A.H.; Chen, A. Mobile augmented reality for teaching structural analysis. *Adv. Eng. Inform.* **2017**, *34*, 90–100. [[CrossRef](#)]
38. Maia, L.F.; Nolêto, C.; Lima, M.; Ferreira, C.; Marinho, C.; Viana, W.; Trinta, F. LAGARTO: A Location based Games Authoring Tool enhanced with augmented reality features. *Entertain. Comput.* **2017**, *22*, 3–13. [[CrossRef](#)]
39. Sekhavat, Y.A.; Parsons, J. The effect of tracking technique on the quality of user experience for augmented reality mobile navigation. *Multimed. Tools Appl.* **2018**, *77*, 11635–11668. [[CrossRef](#)]
40. Chiu, C.-C.; Lee, L.-C. System satisfaction survey for the App to integrate search and augmented reality with geographical information technology. *Microsyst. Technol.* **2018**, *24*, 319–341. [[CrossRef](#)]
41. Gimeno, J.; Portalés, C.; Coma, I.; Fernández, M.; Martínez, B. Combining traditional and indirect augmented reality for indoor crowded environments. A case study on the Casa Batlló museum. *Comput. Graph.* **2017**, *69*, 92–103. [[CrossRef](#)]
42. Brito, P.Q.; Stoyanova, J. Marker versus Markerless Augmented Reality. Which Has More Impact on Users? *Int. J. Hum.-Comput. Interact.* **2018**, *34*, 819–833. [[CrossRef](#)]
43. Rogado, A.B.G.; Quintana, A.M.V.; Mayo, L.L. Evaluation of the Use of Technology to Improve Safety in the Teaching Laboratory. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2017**, *12*, 17–23.
44. Ahn, E.; Lee, S.; Kim, G.J. Real-time Adjustment of Contrast Saliency for Improved Information Visibility in Mobile Augmented Reality. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*; ACM: New York, NY, USA, 2015; p. 199.
45. Léger, É.; Drouin, S.; Collins, D.L.; Popa, T.; Kersten-Oertel, M. Quantifying attention shifts in augmented reality image-guided neurosurgery. *Healthc. Technol. Lett.* **2017**, *4*, 188–192. [[CrossRef](#)]
46. Chu, M.; Matthews, J.; Love, P.E.D. Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study. *Autom. Constr.* **2018**, *85*, 305–316. [[CrossRef](#)]
47. Liu, F.; Seipel, S. Precision study on augmented reality-based visual guidance for facility management tasks. *Autom. Constr.* **2018**, *90*, 79–90. [[CrossRef](#)]
48. Peleg-Adler, R.; Lanir, J.; Korman, M. The effects of aging on the use of handheld augmented reality in a route planning task. *Comput. Hum. Behav.* **2018**, *81*, 52–62. [[CrossRef](#)]
49. Dieck, M.C.T.; Jung, T.H.; Rauschnabel, P.A. Determining visitor engagement through augmented reality at science festivals: An experience economy perspective. *Comput. Hum. Behav.* **2018**, *82*, 44–53. [[CrossRef](#)]
50. Scholz, J.; Duffy, K. We ARE at home: How augmented reality reshapes mobile marketing and consumer-brand relationships. *J. Retail. Consum. Serv.* **2018**, *44*, 11–23. [[CrossRef](#)]
51. Imottesjo, H.; Kain, J.-H. The Urban CoBuilder—A mobile augmented reality tool for crowd-sourced simulation of emergent urban development patterns: Requirements, prototyping and assessment. *Comput. Environ. Urban Syst.* **2018**, *71*, 120–130. [[CrossRef](#)]
52. Barreira, J.; Bessa, M.; Barbosa, L.; Magalhães, L. A Context-Aware Method for Authentically Simulating Outdoors Shadows for Mobile Augmented Reality. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 1223–1231. [[CrossRef](#)] [[PubMed](#)]
53. Fenu, C.; Pittarello, F. Svevo tour: The design and the experimentation of an augmented reality application for engaging visitors of a literary museum. *Int. J. Hum.-Comput. Stud.* **2018**, *114*, 20–35. [[CrossRef](#)]
54. Chiu, C.-C.; Lee, L.-C. Empirical study of the usability and interactivity of an augmented-reality dressing mirror. *Microsyst. Technol.* **2018**, *24*, 4399–4413. [[CrossRef](#)]
55. Michel, T.; Genevès, P.; Fourati, H.; Layaida, N. Attitude estimation for indoor navigation and augmented reality with smartphones. *Pervasive Mob. Comput.* **2018**, *46*, 96–121. [[CrossRef](#)]

56. Torres-Jiménez, E.; Rus-Casas, C.; Dorado, R.; Jiménez-Torres, M. Experiences Using QR Codes for Improving the Teaching-Learning Process in Industrial Engineering Subjects. *IEEE Rev. Iberoam. Technol. Aprendiz.* **2018**, *13*, 56–62. [[CrossRef](#)]
57. Liu, T.-Y.; Tan, T.-H.; Chu, Y.-L. QR code and augmented reality-supported mobile English learning system. In *Mobile Multimedia Processing*; Springer: Berlin, Germany, 2010; pp. 37–52.
58. Fonseca, D.; Martí, N.; Navarro, I.; Redondo, E.; Sánchez, A. Using augmented reality and education platform in architectural visualization: Evaluation of usability and student's level of satisfaction. In Proceedings of the 2012 International Symposium on Computers in Education (SIIE), Andorra la Vella, Andorra, 29–31 October 2012; pp. 1–6.
59. Sánchez, A.; Redondo, E.; Fonseca, D. Developing an Augmented Reality Application in the Framework of Architecture Degree. In Proceedings of the 2012 ACM Workshop on User Experience in e-Learning and Augmented Technologies in Education, New York, NY, USA, 2 November 2012; pp. 37–42.
60. Santana-Mancilla, P.C.; García-Ruiz, M.A.; Acosta-Díaz, R.; Juárez, C.U. Service Oriented Architecture to Support Mexican Secondary Education through Mobile Augmented Reality. *Procedia Comput. Sci.* **2012**, *10*, 721–727. [[CrossRef](#)]
61. Shirazi, A.; Behzadan, A.H. Technology-enhanced learning in construction education using mobile context-aware augmented reality visual simulation. In Proceedings of the 2013 Winter Simulations Conference (WSC), Washington, DC, USA, 8–11 December 2013; pp. 3074–3085.
62. Corrêa, A.G.D.; Tahira, A.; Ribeir, J.B.; Kitamura, R.K.; Inoue, T.Y.; Ficheman, I.K. Development of an interactive book with Augmented Reality for mobile learning. In Proceedings of the 2013 8th Iberian Conference on Information Systems and Technologies (CISTI), Lisboa, Portugal, 19–22 June 2013; pp. 1–7.
63. Ferrer, V.; Perdomo, A.; Rashed-Ali, H.; Fies, C.; Quarles, J. How Does Usability Impact Motivation in Augmented Reality Serious Games for Education? In Proceedings of the 2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), Poole, UK, 11–13 September 2013; pp. 1–8.
64. Redondo, E.; Fonseca, D.; Sánchez, A.; Navarro, I. New Strategies Using Handheld Augmented Reality and Mobile Learning-teaching Methodologies, in Architecture and Building Engineering Degrees. *Procedia Comput. Sci.* **2013**, *25*, 52–61. [[CrossRef](#)]
65. Fonseca, D.; Villagrasa, S.; Valls, F.; Redondo, E.; Climent, A.; Vicent, L. Motivation assessment in engineering students using hybrid technologies for 3D visualization. In Proceedings of the 2014 International Symposium on Computers in Education (SIIE), Logrono, Spain, 12–14 November 2014; pp. 111–116.
66. Sánchez, A.; Redondo, E.; Fonseca, D.; Navarro, I. Academic performance assessment using Augmented Reality in engineering degree course. In Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Madrid, Spain, 22–25 October 2014; pp. 1–7.
67. Fonseca, D.; Villagrasa, S.; Valls, F.; Redondo, E.; Climent, A.; Vicent, L. Engineering teaching methods using hybrid technologies based on the motivation and assessment of student's profiles. In Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Madrid, Spain, 22–25 October 2014; pp. 1–8.
68. Camba, J.; Contero, M.; Salvador-Herranz, G. Desktop vs. mobile: A comparative study of augmented reality systems for engineering visualizations in education. In Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Madrid, Spain, 22–25 October 2014; pp. 1–8.
69. Chen, W. Historical Oslo on a Handheld Device—A Mobile Augmented Reality Application. *Procedia Comput. Sci.* **2014**, *35*, 979–985. [[CrossRef](#)]
70. He, J.; Ren, J.; Zhu, G.; Cai, S.; Chen, G. Mobile-Based AR Application Helps to Promote EFL Children's Vocabulary Study. In Proceedings of the 2014 IEEE 14th International Conference on Advanced Learning Technologies, Athens, Greece, 7–10 July 2014; pp. 431–433.
71. Lai, A.S.Y.; Wong, C.Y.K.; Lo, O.C.H. Applying Augmented Reality Technology to Book Publication Business. In Proceedings of the 2015 IEEE 12th International Conference on e-Business Engineering (ICEBE), Beijing, China, 23–25 October 2015; pp. 281–286.
72. Rogers, K.; Frommel, J.; Breier, L.; Celik, S.; Kramer, H.; Kreidel, S.; Brich, J.; Riemer, V.; Schrader, C. Mobile Augmented Reality as an Orientation Aid: A Scavenger Hunt Prototype. In Proceedings of the 2015 International Conference on Intelligent Environments (IE), Prague, Czech Republic, 15–17 July 2015; pp. 172–175.

73. Jamali, S.S.; Shiratuddin, M.F.; Wong, K.W.; Oskam, C.L. Utilising Mobile-Augmented Reality for Learning Human Anatomy. *Procedia-Soc. Behav. Sci.* **2015**, *197*, 659–668. [CrossRef]
74. Majid, N.A.A.; Mohammed, H.; Sulaiman, R. Students' Perception of Mobile Augmented Reality Applications in Learning Computer Organization. *Procedia-Soc. Behav. Sci.* **2015**, *176*, 111–116. [CrossRef]
75. Zainuddin, N.; Idrus, R.M. The use of augmented reality enhanced flashcards for arabic vocabulary acquisition. In Proceedings of the 2016 13th Learning and Technology Conference (L T), Jeddah, Saudi Arabia, 10–11 April 2016; pp. 1–5.
76. Bazzaza, M.W.; Alzubaidi, M.; Zemerly, M.J.; Weruga, L.; Ng, J. Impact of smart immersive mobile learning in language literacy education. In Proceedings of the 2016 IEEE Global Engineering Education Conference (EDUCON), Abu Dhabi, United Arab Emirates, 10–13 April 2016; pp. 443–447.
77. Qassem, L.M.M.S.A.; Hawaii, H.A.; Shehhi, S.A.; Zemerly, M.J.; Ng, J.W.P. AIR-EDUTECH: Augmented immersive reality (AIR) technology for high school Chemistry education. In Proceedings of the 2016 IEEE Global Engineering Education Conference (EDUCON), Abu Dhabi, United Arab Emirates, 10–13 April 2016; pp. 842–847.
78. Kulpy, A.; Bekaroo, G. Fruitify: Nutritionally augmenting fruits through markerless-based augmented reality. In Proceedings of the 2017 IEEE 4th International Conference on Soft Computing Machine Intelligence (ISCM), Mauritius, Mauritius, 23–24 November 2017; pp. 149–153.
79. Chessa, M.; Solari, F. [POSTER] Walking in Augmented Reality: An Experimental Evaluation by Playing with a Virtual Hopscotch. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 9–13 October 2017; pp. 143–148.
80. Bratitsis, T.; Bardanika, P.; Ioannou, M. Science Education and Augmented Reality Content: The Case of the Water Circle. In Proceedings of the 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), Timisoara, Romania, 3–7 July 2017; pp. 485–489.
81. Selviany, A.; Kaburuan, E.R.; Junaedi, D. User interface model for Indonesian Animal apps to kid using Augmented Reality. In Proceedings of the 2017 International Conference on Orange Technologies (ICOT), Singapore, 8–10 December 2017; pp. 134–138.
82. Müller, L.; Aslan, I.; Krüßen, L. GuideMe: A Mobile Augmented Reality System to Display User Manuals for Home Appliances. In *Advances in Computer Entertainment*; Reidsma, D., Katayose, H., Nijholt, A., Eds.; Springer Nature: Cham, Switzerland, 2013; pp. 152–167.
83. Tsai, T.-H.; Chang, H.-T.; Yu, M.-C.; Chen, H.-T.; Kuo, C.-Y.; Wu, W.-H. Design of a Mobile Augmented Reality Application: An Example of Demonstrated Usability. In *Universal Access in Human-Computer Interaction. Interaction Techniques and Environments*; Springer International Publishing: Cham, Switzerland, 2016; pp. 198–205.
84. Master Journal List—Clarivate Analytics. Available online: <http://mjl.clarivate.com/> (accessed on 17 January 2019).
85. Scimago Journal & Country Rank. Available online: <https://www.scimagojr.com/> (accessed on 17 January 2019).
86. Bhattacharjee, A. *Social Science Research: Principles, Methods, and Practices*; Textbooks Collection; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2012.
87. Abásolo, M.J.; Abreu, J.; Almeida, P.; Silva, T. *Applications and Usability of Interactive Television: 6th Iberoamerican Conference, jAUTI 2017; Aveiro, Portugal, October 12–13, 2017, Revised Selected Papers*; Springer: Berlin, Germany, 2018.
88. Patten, M.L. *Proposing Empirical Research: A Guide to the Fundamentals*; Taylor & Francis: Milton Park, Didcot, UK; Abingdon, UK, 2016.
89. MacKenzie, I.S. *Human-Computer Interaction: An Empirical Research Perspective*; Newnes: Boston, MA, USA, 2012.
90. Niessen, M.; Peschar, J. *Comparative Research on Education: Overview, Strategy and Applications in Eastern and Western Europe*; Elsevier: Amsterdam, The Netherlands, 2013.
91. Randolph, J.J. *Multidisciplinary Methods in Educational Technology Research and Development*; HAMK Press/Justus Randolph: Hameenlinna, Finland, 2008.
92. Purchase, H.C. *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples*; Cambridge University Press: Cambridge, UK, 2012.
93. Srinagesh, K. *The Principles of Experimental Research*; Elsevier: Amsterdam, The Netherlands, 2011.

94. Thyer, B.A. *Quasi-Experimental Research Designs*; Oxford University Press: Oxford, UK, 2012.
95. Thompson, C.B.; Panacek, E.A. Research study designs: Experimental and quasi-experimental. *Air Med. J.* **2006**, *25*, 242–246. [[CrossRef](#)] [[PubMed](#)]
96. Moustakas, C. *Heuristic Research: Design, Methodology, and Applications*; SAGE Publications: Thousand Oaks, CA, USA, 1990.
97. Sultan, N. *Heuristic Inquiry: Researching Human Experience Holistically*; SAGE Publications: Thousand Oaks, CA, USA, 2018.
98. Salhi, S. *Heuristic Search: The Emerging Science of Problem Solving*; Springer: Berlin, Germany, 2017.
99. Salehi, S.; Selamat, A.; Fujita, H. Systematic mapping study on granular computing. *Knowl.-Based Syst.* **2015**, *80*, 78–97. [[CrossRef](#)]
100. ISO, W. 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). *Int. Organ. Stand.* **1998**, *45*, 1–29.
101. Sauro, J.; Lewis, J.R. *Quantifying the User Experience: Practical Statistics for User Research*; Morgan Kaufmann: Burlington, MA, USA, 2016.
102. Wilson, C. *Credible Checklists and Quality Questionnaires: A User-Centered Design Method*; Newnes: Boston, MA, USA, 2013.
103. Portugal, S. *Interviewing Users: How to Uncover Compelling Insights*; Rosenfeld Media: New York, NY, USA, 2013.
104. Isbister, K.; Schaffer, N. *Game Usability: Advancing the Player Experience*; CRC Press: Boca Raton, FL, USA, 2015.
105. James, M.R., Jr.; Dale, M.; James, D.; Minsoo, K. *Measurement and Evaluation in Human Performance, 5E*; Human Kinetics: Champaign, IL, USA, 2015.
106. Bernsen, N.O.; Dybkjær, L. *Multimodal Usability*; Springer Science & Business Media: Berlin, Germany, 2009.
107. Cheng, J.M.-S.; Blankson, C.; Wang, E.S.-T.; Chen, L.S.-L. Consumer attitudes and interactive digital advertising. *Int. J. Advert.* **2009**, *28*, 501–525. [[CrossRef](#)]
108. Lavie, T.; Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.* **2004**, *60*, 269–298. [[CrossRef](#)]
109. Westbrook, R.A.; Oliver, R.L. The Dimensionality of Consumption Emotion Patterns and Consumer Satisfaction. *J. Consum. Res.* **1991**, *18*, 84–91. [[CrossRef](#)]
110. Kizony, R.; Katz, N.; Rand, D.; Weiss, P. Short Feedback Questionnaire (SFQ) to enhance client-centered participation in virtual environments. *Cyberpsychol. Behav.* **2006**, *9*, 687–688.
111. Loureiro, S.M.C. The role of the rural tourism experience economy in place attachment and behavioral intentions. *Int. J. Hosp. Manag.* **2014**, *40*, 1–9. [[CrossRef](#)]
112. Mehmetoglu, M.; Engen, M. Pine and Gilmore's Concept of Experience Economy and Its Dimensions: An Empirical Examination in Tourism. *J. Qual. Assur. Hosp. Tour.* **2011**, *12*, 237–255. [[CrossRef](#)]
113. Oh, H.; Fiore, A.M.; Jeoung, M. Measuring Experience Economy Concepts: Tourism Applications. *J. Travel Res.* **2007**, *46*, 119–132. [[CrossRef](#)]
114. Quadri-Felitti, D.L.; Fiore, A.M. Destination loyalty: Effects of wine tourists' experiences, memories, and satisfaction on intentions. *Tour. Hosp. Res.* **2013**, *13*, 47–62. [[CrossRef](#)]
115. Nielsen, J. *Usability Engineering*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993.
116. Lewis, J.R. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *Int. J. Hum.-Comput. Interact.* **1995**, *7*, 57–78. [[CrossRef](#)]
117. Jordan, P.W.; Thomas, B.; McClelland, I.L.; Weerdmeester, B. *Usability Evaluation in Industry*; CRC Press: Boca Raton, FL, USA, 1996.
118. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [[CrossRef](#)]
119. Pintrich, P.R.; Smith, D.A.F.; Garcia, T.; Mckeachie, W.J. Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educ. Psychol. Meas.* **1993**, *53*, 801–813. [[CrossRef](#)]
120. Olsson, T. Concepts and subjective measures for evaluating user experience of mobile augmented reality services. In *Human Factors in Augmented Reality Environments*; Springer: Berlin, Germany, 2013; pp. 203–232.
121. Lewis, J.R. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **1992**, *36*, 1259–1260. [[CrossRef](#)]
122. Pifarré, M.; Tomico, O. Bipolar Laddering (BLA): A Participatory Subjective Exploration Method on User Experience. In Proceedings of the 2007 Conference on Designing for User eXperiences, New York, NY, USA, 5–7 November 2007; p. 2.

123. Martín-Gutiérrez, J.; Saorín, J.L.; Contero, M.; Alcañiz, M.; Pérez-López, D.C.; Ortega, M. Design and validation of an augmented book for spatial abilities development in engineering students. *Comput. Graph.* **2010**, *34*, 77–91. [[CrossRef](#)]
124. O'Brien, H.L.; Toms, E.G. The development and evaluation of a survey to measure user engagement. *J. Am. Soc. Inf. Sci. Technol.* **2010**. Available online: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.21229> (accessed on 8 January 2019).
125. Padda, H. *QUIM: A Model for Usability/Quality in Use Measurement*; LAP LAMBERT Academic Publishing: Saarbrücken, Germany, 2009.
126. Venkatesh, V.; Thong, J.Y.L.; Xu, X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Q.* **2012**, *36*, 157–178. [[CrossRef](#)]
127. Buchanan, T.; Paine, C.; Joinson, A.N.; Reips, U.-D. Development of measures of online privacy concern and protection for use on the Internet. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 157–165. [[CrossRef](#)]
128. Mathwick, C.; Malhotra, N.; Rigdon, E. Experiential value: Conceptualization, measurement and application in the catalog and Internet shopping environment☆11☆This article is based upon the first author's doctoral dissertation completed while at Georgia Institute of Technology. *J. Retail.* **2001**, *77*, 39–56. [[CrossRef](#)]
129. Cockton, G.; Lavery, D.; Woolrych, A. *The Human-computer Interaction Handbook*; Jacko, J.A., Sears, A., Eds.; Lawrence Erlbaum Associates Inc.: Hillsdale, NJ, USA, 2003; pp. 1118–1138.
130. Abellán, I.M.J.; Stake, R.E. Does Ubiquitous Learning Call for Ubiquitous Forms of Formal Evaluation? An Evaluand oriented Responsive Evaluation Model. *Ubiquitous Learn. Int. J.* **2009**, *1*, 71–82.
131. Nielsen, J. Usability Inspection Methods. In *Conference Companion on Human Factors in Computing Systems*; ACM: New York, NY, USA, 1994; pp. 413–414.
132. Ko, S.M.; Chang, W.S.; Ji, Y.G. Usability Principles for Augmented Reality Applications in a Smartphone Environment. *Int. J. Hum.-Comput. Interact.* **2013**, *29*, 501–515. [[CrossRef](#)]
133. Gómez, R.Y.; Caballero, D.C.; Sevillano, J.-L. Heuristic Evaluation on Mobile Interfaces: A New Checklist. *Sci. World J.* **2014**. Available online: <https://www.hindawi.com/journals/tswj/2014/434326/> (accessed on 9 January 2019).
134. Dix, A. *Human-computer interaction*; Springer: Berlin, Germany, 2009.
135. Dünser, A.; Billinghamurst, M. Evaluating augmented reality systems. In *Handbook of Augmented Reality*; Springer: Berlin, Germany, 2011; pp. 289–307.
136. Lim, K.C. *A Comparative Study of GUI and TUI for Computer Aided Learning*; Universiti Tenaga Nasional: Selangor, Malaysia, 2011.
137. Uras, S.; Ardu, D.; Paddeu, G.; Deriu, M. Do Not Judge an Interactive Book by Its Cover: A Field Research. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*, New York, NY, USA, 3–5 December 2012; pp. 17–20.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).