Deanship of Graduate Studies
Al-Quds University

Framework for Augmented Reality User Experience: Automated Measures of User Experience and Associated Factors

Balqees Ahmad Khaleel Awawdeh

M.Sc. Thesis

Jerusalem-Palestine

1446 / 2025

Framework for Augmented Reality User Experience: Automated Measures of User Experience and Associated Factors

Prepared By:

Balqees Ahmad Khaleel Awawdeh

M.Sc. Thesis

Supervisor: Dr. Radwan Qasrawi

A Thesis Submitted in Partial Fulfillment of Requirements for The Degree of Master in Computer Science / Department of Computer Science / Faculty of Graduate Students / Deanship of Graduate Studies / Al-Quds University.

1446 / 2025

Deanship of Graduate Studies
Al-Quds University

Computer Science

Thesis Approval

Framework for Augmented Reality User Experience: Automated Measures of User Experience and Associated Factors

Prepared By: Balqees Ahmad Khaleel Awawdeh
Registration No: 21920128

Supervisor: Dr. Radwan Qasrawi

Master thesis submitted and accepted, Date: 12/01/2025

1- Head of Committee: Dr.Radwan Qasrawi
2- Internal Examiner: Dr.Nael Abu Al Halaweh
3- External Examiner: Dr.Sobhi Ahmad

Jerusalem – Palestine

1446 / 2025

Dedication

الى من وهبوني الحياة والامل، وعلموني أن ارتقي سلمها بحكمة وصبر ووفاء لهما، والديّ العزيزين

إلى سندي وقوتي في هذه الحياة تعبيرًا عن المحبة العميقة والامتنان الكبير، اخوتي الاعزاء

إلى من كاتفتني ونحن نشق العمر معا وننهل من ربيع أيامه، أختي الطيبة

الى الايادي الصغيرة التي تطرق بابي زائرة أبناء إخوتي الأحباء

الى الأصدقاء الأوفياء الداعمين في أحلك الظروف

الى أولئك اللذين يفرحهم نجاحنا وتألقنا

إليهم جميعاً.. أهدي هذا البحث

Declaration

I certify that this thesis submitted for the degree of Master, is the result of my research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed:

Balqees Ahmad Khaleel Awawdeh

Date: 12/01/2025

Acknowledgment

I would like to express my sincere gratitude to my advisor Dr. Radwan Qasrawi who made this work possible. Many thanks for his guidance, advice, patience, and knowledge.

I would also like to thank my thesis committee.

I would also like to thank my family for continuous support and understanding when undertaking my research. Your prayer for me was what sustained me so far.

Abstract

Background: Augmented reality (AR) technology is rapidly advancing, offering new possibilities for user experience (UX) across fields such as education, tourism, and architecture. Nevertheless, the development of automated frameworks that thoroughly evaluate UX is crucial, particularly in terms of learnability, efficiency, effectiveness, and memorability, based on established human-centered interaction standards.

Objective: This study aims to develop a framework for testing user experience in AR applications, with a primary focus on evaluating usability and user engagement metrics. The framework uses established Human-Computer Interaction principles as a reference for automated measures, providing a comprehensive and objective assessment tool to support the development of intuitive, user-friendly AR interfaces that enhance user engagement.

Method: The automated framework was applied to assess usability metrics across three AR applications—Farah App (educational), Dar Al Consul (tourism), and EasyApp (architectural)—involving a group of 20 users across different age groups. Data on user interactions were collected and analyzed, while Participant responses was gathered through questionnaires to complement the automated analysis.

Results: Findings revealed notable differences in ease of learning, effectiveness, and memorability among the applications. Farah App showed particularly high learnability, especially among younger users, making it a promising educational tool for that demographic. Both Dar Al Consul and EasyApp demonstrated consistent usability across age groups, though minor improvements in task flow and error reduction could further enhance user efficiency.

Conclusion: This study highlights the importance of using HCI-based frameworks to evaluate AR user experience. The developed framework provides comprehensive and objective measurements that support enhanced UX by offering data-driven insights for more intuitive design. The results indicate that AR developers should focus on age-appropriate interactions, streamlined task flows, and efficient feedback mechanisms to improve user engagement and ensure effective, sustainable experiences across a range of AR applications.

Table of Contents

List of Tables

List of Figures

List of acronyms

Acronym	**Meaning**
HCI	Human Computer Interaction
UX	User Experience
VARTOOL	Virtual Auto Augmented Reality Tool
AR	Augmented Reality
VR	Virtual Reality
UI	User Interface
ISO	International Organization for Standardization
DB	Database
DAC	Dar Al Consul
STF	Satisfaction
GUI	Graphical User Interface

Chapter 1:

Introduction

1.1 Introduction

Learning across scientific and practical fields is essential for advancing knowledge and fostering innovation. However, researchers and practitioners often encounter barriers when exploring novel methodologies and advanced technologies, particularly in high-risk or complex domains. Virtual models have become instrumental in overcoming these barriers, enabling the simulation of real-world phenomena and processes to facilitate experimentation and training without direct exposure to risks. Among virtual technologies, augmented reality (AR) has gained significant traction in recent years, providing users with immersive, interactive experiences by overlaying digital information onto the physical world [1]. As AR technology matures, it offers unique learning environments where users can engage dynamically with content, thus fostering higher levels of engagement and comprehension than traditional teaching methods [2].

Despite its promise, AR development faces numerous challenges, particularly in ensuring an optimal user experience (UX) and effective human-computer interaction (HCI) design. As a relatively new medium, AR lacks well-defined design principles tailored to its unique characteristics, which are essential for creating user-centered applications. This gap presents significant obstacles for developers seeking to build effective AR solutions, as user experience directly influences usability, learnability, and the overall acceptance of AR technologies [3]. [3]In immersive applications like AR, learnability—the ease with which new users can accomplish basic tasks—plays a critical role in determining whether users will adopt or abandon the technology. As highlighted by ISO standards, learnability is a key measure of system usability, emphasizing that users should be able to complete tasks effectively, efficiently, and with satisfaction [4].[4]

The potential of AR as an educational tool is reflected in its growing adoption across sectors, including education, healthcare, architecture, and tourism [5]. By offering engaging, hands-on experiences, AR has been shown to improve learning outcomes and retention rates. For instance, Rasheed et al. (2021) demonstrated the benefits of a virtual laboratory developed with AR principles, showing improved learning outcomes among students in a desktop-based environment compared to those in traditional physical labs. However, ensuring AR applications are easy to learn remains a challenge, as conventional usability and UX evaluation methods, such as task-based testing and heuristic evaluations, are often time-consuming, labor-intensive, and expensive [6]. These methods require extensive human resources and introduce the possibility of human biases, underscoring the need for innovative, automated solutions that can streamline evaluation processes without sacrificing accuracy.

Integrating systematic UX evaluation tools in AR product development is essential to visualize and enhance the conceptual user experience effectively. Automated evaluations reduce time and costs while providing developers with actionable insights into user interaction patterns, learnability factors, and engagement levels [7]. Drawing on established usability principles, this research aims to develop a comprehensive, automated framework to evaluate AR learnability, thereby empowering developers to refine AR applications for improved user experience, increased engagement, and efficient learning outcomes. This study will focus on providing developers with robust, automated tools to measure and enhance user experience, ultimately supporting the usability and adoption of AR technologies across diverse applications.

1.2 Problem Statement

Augmented Reality (AR) is revolutionizing multiple industries by merging digital and real-world environments, creating immersive experiences that enhance user engagement and operational efficiency. AR's capacity to offer interactive, context-aware experiences has made it invaluable in domains such as education, healthcare, entertainment, and tourism, where it enhances experiential learning and provides real-time information overlays ([8],[9]) However, the effectiveness of AR applications depends largely on their usability and the quality of user experience, which are critical for user engagement and satisfaction. Adhering to HCI standards, including principles like user-centered design, ease of use, and overall satisfaction, is essential to ensuring AR applications are effective and accessible [10].

Traditional usability testing methods for AR and VR technologies, while well-established, are often insufficient for deeply interactive experiences. Task-based testing, think-aloud protocols, heuristic evaluations, and post-experience questionnaire have been widely used, but these methods can be time-consuming, labor-intensive, and cost-prohibitive, often requiring extensive human involvement and repeated testing. This limits their scalability and introduces potential biases due to human subjectivity [11]. Furthermore, existing automated usability tools are often designed for standard graphical user interfaces (GUIs) and fail to address the specific learnability challenges posed by AR applications, which rely on dynamic interactions and immediate responsiveness. Current automated tools also frequently lack accessibility for non-English-speaking users and present integration challenges with application source codes, making them less adaptable for global and cross-platform applications[12].

Learnability, defined as the speed and ease with which users accomplish tasks upon first use, is a pivotal factor in AR's adoption and success. High learnability contributes to better user performance, sustained engagement, and successful long-term adoption [13]. However, neither traditional nor automated testing methods adequately capture learnability metrics specific to AR's interactive, immersive environments. As a result, developers lack comprehensive tools to evaluate and optimize user experience metrics in AR, particularly in widely accessible applications that cater to diverse user bases.

This research aims to address these limitations by proposing a novel, automated framework for evaluating AR UX metrics as learnability, effectiveness, efficiency and memorability. By integrating human-based evaluation methodologies with automated assessment tools, this framework offers a scalable solution that aligns with HCI standards and enhances the efficiency and effectiveness of AR usability testing. This innovative framework is expected to facilitate the development of user-friendly AR applications that support user engagement and meet global UX standards, providing developers with a reliable tool to improve learnability and usability across diverse AR platforms.

1.3 Research Questions

1. Does the proposed automated framework improve the efficiency and effectiveness of user experience testing for AR applications by reducing the time and resources required?
2. How accurately does the framework assess UX in AR applications, specifically concerning users' ability to complete tasks?
3. To what extent does the framework contribute to enhancing the overall user experience (UX) quality of AR applications?

1.4 Objectives of the Study

The primary objective of this research is to develop a practical, automated framework for evaluating the user experience of AR applications, with the goal of enhancing associated factors and engagement. By implementing automated measures, this framework aims to streamline the AR evaluation process, reduce the time and resources required, and provide developers with a comprehensive tool for optimizing AR applications. Key objectives include:

1. Identifying the key factors that affect AR usability, including learnability, efficiency, effectiveness, and memorability, analyzing their impact on the overall user experience.
2. Developing structured, automated measures of user experience and related factors for AR applications, integrating these elements to create a robust evaluation system.
3. Implementing the framework to assess user experience components, facilitating a more efficient, accurate, and scalable evaluation of AR applications.
4. Testing the framework's effectiveness through experiments, ensuring it reliably measures Learnability for AR applications.
5. Providing actionable insights to improve AR learnability based on empirical results, aiming to enhance UX and engagement across different AR applications.

1.5 Thesis Contribution

This thesis makes several significant contributions to the field of AR usability and learnability evaluation. The primary contributions are:

1. The development of an automated framework for assessing the learnability of AR applications, grounded in HCI principles, which offers developers an efficient tool for evaluating and improving UX in AR applications.
2. Increased efficiency in time, cost, and effort required for learnability testing through automation, providing developers with a scalable solution to optimize AR applications for diverse user groups.
3. Introduction of a comprehensive learnability evaluation framework that integrates HCI standards and automated tools, facilitating a structured, efficient evaluation of UX in AR.

1.6 Research Methodology

This study proposes an automated tool based on a structured framework aligned with HCI standards, specifically using established UX principles, to assess AR application learnability. The evaluation outcomes are presented in multiple formats, including tabular views, dashboards, and comparative analyses, enabling developers to interpret findings easily.

Step 1: Establishing HCI Standards

Nielsen's UX principles were adapted to define specific HCI measures that serve as benchmarks for AR user experience evaluation. These principles were operationalized into quantifiable criteria with specific cut-off values, stored in a database accessible to AR developers for comparative evaluations.

Step 2: Designing the Platform and Dashboard

The platform for data collection and visualization was built using Unity, Python, HTML, JavaScript, and jQuery, and utilized Firebase for data storage. Three separate databases were created per application, ensuring reliable and efficient data storage and retrieval.

Step 3: Data Collection

Data collection involved defining specific tasks within Unity-based AR applications. Selected samples engaged with these tasks via mobile devices, with interaction data stored in Firebase in a structured format for easy analysis.

Step 4: Data Analysis and Visualization

The collected data was exported as JSON files and visualized in dashboards and tables. The real-time Firebase database enabled dynamic updates, ensuring timely and accurate learnability analysis across different AR applications.

1.7 Research Limitations

A primary limitation encountered was the scarcity of AR open-source code for testing the developed tool, which led to reliance on locally sourced applications from various fields.

1.8 Thesis Organization

Chapter 1 introduces the study, presenting the problem statement, research questions, objectives, methodology, limitations, and contributions.
Chapter 2 reviews background topics including HCI UX, AR environments, the Unity platform, Firebase, ISO standards, and usability testing methods, alongside a literature review of UX evaluation in AR.
Chapter 3 details the methodology, covering the development of the automated framework, VARTOOL design, and testing/evaluation processes.
Chapter 4 presents research results, discussing tool performance and effectiveness.

Chapter 5 discusses key findings, summarizes the research, and suggests areas for future work.

Chapter 2:

Background and Literature Review

2.1 Background

The increasing integration of augmented reality (AR) into diverse sectors such as education, tourism, and healthcare underscore the need for a comprehensive understanding of its foundational principles and associated user experience (UX) challenges. AR environments merge virtual and physical elements to create immersive experiences, but their usability and learnability are critical to user adoption and sustained engagement. In recent literature, researchers have emphasized the importance of human-computer interaction (HCI) principles in designing intuitive AR systems that reduce cognitive load and improve interaction quality. Studies highlight that frameworks for assessing AR UX remain underdeveloped, particularly in areas such as automated evaluation of learnability, efficiency, and effectiveness. The growing body of research in HCI and AR provides a valuable foundation for addressing these gaps, offering insights into methods for improving AR applications through systematic design and iterative testing. By synthesizing findings from prior studies, this chapter establishes the groundwork for exploring innovative frameworks to enhance the AR user experience.

2.1.1 Human-Computer Interaction (HCI) and User Experience (UX)

Human-Computer Interaction (HCI) is a multidisciplinary field dedicated to the design, evaluation, and implementation of interactive computing systems for human use, with a focus on understanding the interactions between humans and machines. Rooted in psychology, computer science, design, and engineering, HCI seeks to improve the usability, accessibility, and overall user satisfaction of computing systems [14]. At its core, HCI emphasizes the need for systems that align with human cognitive and emotional processes, ensuring that interactions are intuitive and engaging [15].

Within HCI, User Experience (UX) has emerged as a critical dimension, focusing not only on functionality but also on the emotional and psychological effects of user interactions with a system. UX encompasses a wide range of factors, including usability, accessibility, aesthetics, and overall satisfaction, each contributing to the user's perception and emotional response to the system [16] Unlike traditional usability, which prioritizes task efficiency and error reduction, UX considers the broader context of the interaction, taking into account user expectations, cultural factors, and personal preferences [17]. Thus, UX is inherently subjective, influenced by a user’s emotions, perceptions, and prior experiences, as well as the context in which the interaction takes place.

The quality of UX has been shown to significantly impact user behavior and decision-making. Positive UX can lead to increased user engagement, higher retention rates, and positive word-of-mouth, while negative UX can result in frustration, reduced productivity, and ultimately the abandonment of the system [18]. This is particularly important in emerging technologies, such as augmented reality (AR), where intuitive and satisfying interactions are vital for user acceptance and sustained engagement [19]. Within Human-Computer Interaction (HCI), various methodologies like usability testing, heuristic evaluation, and cognitive walkthroughs have been established to assess and enhance user experience (UX). These methods empower designers and developers to uncover potential usability problems, understand user preferences, and refine systems based on user feedback and empirical data. [20].

In the context of AR, UX must be adapted to account for the unique challenges presented by interactive, real-time interactions with digital objects in physical spaces. Key UX dimensions such as usability, performance, ease of use, and user enjoyment are crucial for AR applications, as they determine not only how users interact with AR systems but also how they perceive and engage with the digital content overlaid on the real world [3]. Therefore, a comprehensive understanding of HCI and UX principles is essential for creating AR experiences that are not only functional but also emotionally engaging and user-centered.

2.1.2 Augmented Reality (AR) Environment

Augmented Reality (AR) is a technology that integrates computer-generated elements with the real-world environment, creating immersive, interactive experiences where users can interact with both physical and digital objects simultaneously [1]. Unlike Virtual Reality (VR), which immerses users in a completely virtual world, AR overlays digital content onto real-world scenes, enhancing the user's perception and interaction within their actual environment. AR environments are designed to provide real-time interaction and spatial alignment between virtual and physical elements, facilitated by devices like smartphones, tablets, and AR glasses [9].

The fundamental components of AR include hardware, software, and user interface (UI) design. **Hardware** includes the devices used to deliver AR experiences, such as smartphones equipped with cameras, accelerometers, and gyroscopes, as well as head-mounted displays that provide a more immersive experience [21]. **Software** encompasses the platforms and development tools that support AR functionality, such as Unity, ARKit, and ARCore, which offer frameworks for detecting and interacting with physical spaces in real time [22]. **UI Design** in AR is distinct from traditional interfaces, as it must consider both the physical and digital dimensions of interaction, ensuring that the interface is intuitive and non-intrusive [23]. In particular, accurate registration—the precise alignment of digital elements within the physical environment—is critical for maintaining the illusion of coexistence between virtual and real-world objects [8].

AR environments have demonstrated considerable potential in applications such as education, healthcare, and industrial training by providing interactive, contextually relevant content that enhances user engagement and comprehension. For example, AR can enable medical students to view anatomical structures overlaid on physical models, offering a hands-on learning experience that improves understanding and retention [24]. Similarly, AR has been shown to facilitate training in industrial contexts, where real-time, step-by-step instructions can be displayed directly within a technician's field of view, reducing error rates and improving efficiency [25]. The immersive nature of AR environments makes them ideal for UX and usability testing, allowing researchers to observe real-time user responses and behaviors in a controlled yet realistic setting [26].

2.1.3 User Experience of Augmented Reality Applications

The user experience (UX) of AR applications is complex, as it involves interactions that are more immersive and multi-dimensional compared to traditional applications. Effective UX in AR is crucial for ensuring user satisfaction, sustained engagement, and long-term adoption of AR technologies [27]. Due to the mixed-reality nature of AR, UX must account for user expectations of both physical and virtual elements, providing a seamless experience where digital content is perceived as an extension of the real world [9]. This is especially important as AR applications are increasingly used in critical domains such as education, healthcare, and industrial design, where effective UX can directly impact learning outcomes, patient care, and operational efficiency [28].

A significant factor in the UX of AR applications is learnability—how easily users can understand and effectively use the technology upon first interaction. Learnability influences user satisfaction, productivity, and overall engagement with AR applications, impacting adoption rates and success in real-world applications In many cases, users must learn not only how to interact with the digital interface but also how to interpret and manipulate virtual elements within a physical space. The dual task of navigating both virtual and physical environments present unique challenges, as users must cognitively adapt to a new interaction paradigm that is both spatially and contextually complex [21].

Dunleavy and Dede (2014) showed that traditional methods of evaluating UX, such as user surveys, think-aloud protocols, and interviews, have limitations in AR contexts, as they often fail to capture the nuances of spatial interaction and cognitive load associated with AR environments [29]. As a result, there is a need for a comprehensive, scalable framework that evaluates AR-specific UX factors, including the ease of understanding spatial interactions and adapting to immersive environments. A structured evaluation framework for AR UX would allow developers to systematically assess and optimize user engagement, satisfaction, and learnability ultimately enhancing the usability and appeal of AR applications [30].

2.1.4 Importance of Learnability in AR User Experience

Learnability is a pivotal aspect of UX in AR applications, as it determines how quickly and efficiently users can perform tasks within an AR environment. High learnability facilitates smoother interactions, reduces cognitive load, and increases user satisfaction, leading to a more positive initial experience that is essential for user retention and long-term engagement [31]. In contrast, low learnability can result in user frustration, increased error rates, and ultimately, abandonment of the application [32]. This is particularly relevant in AR, where complex interactions with virtual elements in physical spaces can significantly increase cognitive demands on the user [33].

Empirical studies have shown that learnability is closely linked to user satisfaction and engagement in technology-driven interactions. Users are more likely to engage with applications that are easy to learn and intuitively designed, and they are more likely to abandon applications that have a steep learning curve [34] .In AR applications, learnability is impacted by multiple factors, including the intuitiveness of the user interface, the clarity of virtual-to-real-world interactions, and the user's ability to understand spatially anchored digital content. Improving learnability in AR requires a comprehensive understanding of these factors and the development of design principles that reduce the learning curve for first-time users [16].

2.1.5 Automated Measures for Evaluating Learnability

Traditional methods for evaluating learnability, such as direct observation and user surveys, provide valuable insights but are resource-demanding and difficult to apply on a large scale [35]. With the advent of automated testing tools, researchers can now assess learnability more efficiently and objectively. Automated measures offer several advantages, including real-time data collection, reduced human bias, and the ability to capture implicit metrics such as task success rates, learning curves, and user errors [36]. In AR, automated evaluation tools can track user interactions, analyze cognitive load through eye-tracking data, and monitor task completion times, providing a more comprehensive picture of user learning behaviors [33].

Automated evaluation methods can be particularly useful in AR, where interactions are more complex and multidimensional compared to traditional applications. For instance, automated tools can record user experience measures, such as task success rates and error rate, which are then analyzed to assess learnability across different stages of user interaction. Integrating advanced analytics, such as machine learning algorithms, into these tools allows them to adapt to individual learning patterns, thereby providing developers with actionable insights that can inform iterative improvements to AR systems [9].

2.1.6 Key Factors Affecting Learnability in Augmented Reality

Learnability in AR is influenced by a range of factors, including individual user traits, contextual factors, and the design of the system. Cognitive abilities, prior experience with technology, and familiarity with AR interfaces play a significant role in how quickly users can adapt to new AR systems [24]. Environmental factors, such as lighting, noise, and the accuracy of AR tracking, also impact learnability, as they affect the clarity and stability of the digital content overlay, influencing user interactions and task efficiency [8].

System design is another critical factor. Well-designed interfaces with intuitive controls, clear visual cues, and immediate feedback mechanisms are more likely to enhance learnability, while complex or cluttered interfaces may hinder it [35]. Additionally, cognitive and emotional factors, such as motivation, engagement, and attentional focus, play a significant role in determining how quickly and effectively users can learn to use AR systems [33]. Interfaces that engage users cognitively and emotionally support faster learning and improve long-term retention, emphasizing the importance of user-centered design in AR environments [36].

2.2 Literature Review

2.2.1 UX Standards in AR Application Design

To create effective and engaging Augmented Reality (AR) applications, adherence to User Experience (UX) standards is essential. UX design principles in AR not only enhance usability but also help foster user satisfaction, safety, and engagement. Panagiotis (2011) introduced a comprehensive UX framework specific to AR and Virtual Reality (VR), emphasizing core elements such as the sense of presence, ergonomics, health and safety, overall usability and product identification. This framework highlights that AR, while a powerful tool for immersive experiences, demands careful adherence to UX principles to ensure intuitive, safe, and meaningful interactions [37]. The application of UX standards is particularly crucial in AR as it integrates both digital and physical worlds, where any design inconsistencies or errors can disrupt the user experience and diminish the perceived value of the application.

Jerald (2015) provides further insights into UX by exploring the design elements of VR systems, with key principles such as immersion, presence, and interaction design that are equally applicable to AR environments. In both VR and AR, immersion refers to the extent to which the user feels present in an augmented or virtual environment, and this sense of immersion plays a crucial role in enhancing user engagement and overall satisfaction. Jerald emphasizes that successful AR experiences require high levels of immersion and seamless interaction, where users can intuitively engage with virtual objects without disruptions or discomfort [38].Such standards are critical in creating AR applications that not only captivate users but also encourage extended interaction, which is a benchmark for user acceptance and satisfaction.

International standards, such as ISO/IEC 25010:2011, provide a software quality framework that includes usability, efficiency, and satisfaction—factors central to designing and evaluating AR/VR applications [39]. These guidelines help ensure that AR applications meet global quality standards, supporting scalability and user trust. Cipresso et al. (2018) reinforce the importance of UX-driven design in AR, emphasizing that user-friendly and highly interactive interfaces significantly enhance user experience and minimize cognitive overload. Their research underscores that AR applications must prioritize UX in every stage of development to provide seamless navigation, effective visual cues, and engaging interactions, all of which contribute to a positive user experience [25].

Steed and Oliveira (2009) examine the specific challenges and solutions in designing networked AR environments, particularly in collaborative settings. Their study addresses the need for synchronized interactions, where multiple users can interact with the same virtual objects in real time. They argue that UX principles in AR should extend to facilitating smooth, real-time collaboration, which adds complexity but also enhances the user experience by making AR applications suitable for social and collaborative tasks [40]. In this context, networked environments and collaborative features are increasingly relevant as AR becomes a tool for team-based projects in fields like engineering, healthcare, and remote training.

Usability engineering methodologies are also integral are also integral to AR UX design, as demonstrated by Gabbard and Swan (2008), who emphasize that user-centered evaluations provide critical insights for refining AR applications. Their work highlights the importance of testing AR applications with real users, as this enables designers to gather feedback on specific usability issues and adjust the design to improve user experience. This feedback loop is essential in creating AR applications that respond to user needs, ensuring that the applications are not only technically robust but also aligned with user expectations [41].

Bowman and McMahan (2007) explore immersion's impact on UX by investigating how various immersion levels affect user performance and satisfaction in VR environments. Their findings are relevant to AR, where a high degree of immersion can enhance user engagement. However, they caution that too much immersion without proper UX design can lead to disorientation or discomfort, emphasizing the need for a balance between immersive experiences and functional user interaction in AR [42].

Presence is another critical element of UX in AR, as it directly impacts user engagement. Schuemie et al. (2001) provide a foundational analysis of the factors that contribute to the perception of being immersed in the environment in AR/VR environments, such as sensory fidelity and interaction quality. Presence, or the feeling of being “inside” the augmented environment, is essential for creating an engaging AR experience that feels real to the user. High presence enhances users' willingness to interact with virtual objects as though they exist in the real world, which is particularly beneficial in applications for training, education, and therapy [43].

Norman’s (2013) *The Design of Everyday Things* contributes a set of design principles that emphasize ease of use, intuitiveness, and functionality, which are applicable in AR UX design. Norman's concept of "affordances," where the design suggests possible actions, is particularly relevant in AR applications, where users need immediate understanding of how to interact with virtual objects layered over the real world. His principles underscore the importance of intuitive design, especially for first-time AR users who may be unfamiliar with AR interaction mechanics [17].

The importance of UX in AR design is also evident in Hassenzahl and Tractinsky’s (2006) work on UX within Human-Computer Interaction (HCI). They highlight the affective and sensory dimensions in UX, which are just as important as usability, arguing that positive user experiences depend on deep user connection and visual attractiveness. Their research is critical for AR design, as AR applications that lack engaging and visually appealing elements may struggle to retain users, regardless of technical functionality [16]. Jakob Nielsen’s seminal work Usability Engineering (1993) introduces practical methods for testing and evaluating UX that are relevant to AR, such as heuristic evaluations, which can be used to iteratively refine AR interfaces and address usability issues early in the development process.

ISO 9241-210:2010 offers guidelines for human-centered design that are critical for AR, emphasizing that UX design should prioritize ergonomics and adaptability to diverse user needs (ISO 9241-210, 2010). Garrett’s (2010) layered UX model, which categorizes UX into strategy, scope, structure, skeleton, and surface, provides a systematic approach to designing AR applications. This model enables developers to address UX at multiple levels, from conceptual design to the final interface, ensuring a holistic and user-centered design process [44]. In *Designing the User Interface*, Shneiderman et al. (2017) advocate for effective HCI strategies that enhance

UX through thoughtful interface design. Their emphasis on simplicity and clarity is particularly relevant in AR, where overloaded interfaces can quickly overwhelm users and reduce overall UX quality.

2.2.2 AR-Supported Product Design

AR has become a revolutionary tool in product design, allowing designers to visualize, prototype, and interact with digital models in immersive environments. Schumann (2020) presents a thorough analysis of AR's role in product design, adapting the Double Diamond design process model to guide AR use cases. The Double Diamond model, which includes Discover, Define, Develop, and Deliver phases, provides a structured methodology for integrating AR into the design workflow. This approach enables designers to explore multiple design solutions before committing to physical prototypes, reducing costs and time [45].

Gao, Zhang, and Li (2018) offer a broad review of AR applications in various stages of product design, highlighting both its benefits and challenges. Their study notes that AR enhances visualization capabilities and facilitates interactive prototyping, which accelerates design iterations. However, they also acknowledge challenges, including hardware limitations and the need for seamless data integration, which can hinder the scalability of AR in design processes. AR’s capacity to overlay digital information on physical objects allows for real-time modifications, enabling designers to visualize complex design elements in ways traditional methods cannot [46].

Building on these insights, Zhang and Wang (2019) propose a detailed framework for incorporating AR into product design, emphasizing how real-time feedback enhances design creativity and efficiency. Their framework allows designers to experiment with and refine prototypes more intuitively, which is particularly valuable in sectors that require rapid innovation and precision. AR’s interactive elements also foster collaboration by enabling multiple stakeholders to engage with digital prototypes simultaneously, improving communication and reducing misunderstandings [46].

Pérez and Mendoza (2021) further illustrate the impact of AR on the prototyping process by highlighting its ability to improve accuracy and speed. Their research found that AR allows designers to detect design flaws early in the process, reducing the need for extensive revisions later on. This efficiency is particularly valuable in industries like automotive and aerospace, where minor errors can lead to significant downstream costs. By providing a three-dimensional, manipulable view of the product, AR enables designers to experiment and refine their ideas in a virtual space that closely mirrors the physical product [47].

Kim and Lee (2017) conducted an empirical study exploring AR’s role in enhancing product design from a user-centered perspective. They found that AR facilitates better understanding of how end-users will interact with the product by enabling designers to visualize product functionality in context. This user-centered approach is particularly important in consumer goods, where understanding user interaction patterns can lead to more intuitive, user-friendly designs [48].

2.2.3 Integration of User-Centered Design in AR Systems

The integration of User-Centered Design (UCD) principles is critical in creating intuitive and effective AR systems. Duenser (2007) offers a foundational review of UCD principles in AR, emphasizing the interdisciplinary collaboration necessary to develop AR-specific guidelines that prioritize user needs. His work highlights the importance of drawing from HCI, cognitive psychology, and usability engineering to create comprehensive design frameworks that meet user expectations and functional requirements [49].

Henderson and Feiner (2011) provide further insights into how AR can support UCD processes, particularly in visualizing user interactions. Their study shows that AR enables designers to observe user interactions with digital objects in a real-world context, facilitating design decisions that are more responsive to user needs. By enabling real-time adjustments, AR supports a design process that is iterative and highly adaptable to user feedback, resulting in applications that align closely with user expectations [50].

Guerra-Filho and Gagliardi (2020) discuss the challenges of adapting traditional UCD methodologies to AR, which often involves spatial and interactive complexities. They propose solutions like iterative testing, user feedback loops, and ergonomic considerations to refine AR interface design, addressing challenges such as limited field-of-view and interface complexity. Their work provides a practical roadmap for incorporating UCD principles into AR applications, ensuring that user interactions are both intuitive and satisfying [51].

Brusilovsky and Millán (2007) focus on the broader application of UCD in AR, emphasizing early and continuous user involvement to ensure that AR systems meet user needs. They argue that iterative testing with real users is essential for refining AR interfaces, as it allows developers to gather valuable insights into usability and adjust the system accordingly. This approach aligns with UCD’s core tenet of prioritizing user needs throughout the development cycle [52].

Baird and McGonigal (2018) present a case study that illustrates the practical application of UCD in AR. Their study details the iterative design and development of an AR application, demonstrating how UCD principles such as usability testing, feedback sessions, and prototype refinement were applied to enhance user satisfaction. The study highlights the value of UCD in creating AR applications that are both functional and enjoyable, proving that continuous user involvement leads to better UX outcomes [53].

2.2.4 AR in Educational Contexts

AR has shown significant potential to transform educational experiences by providing immersive, interactive learning environments. Mena-Vargas (2019) conducted a pivotal study on AR’s role in interdisciplinary learning, demonstrating that AR not only fosters curiosity and engagement but also facilitates effective learning processes across various subjects. His findings suggest that AR can bridge multiple disciplines, supporting innovative educational strategies that cater to diverse learning styles [54].

Ibáñez and Delgado-Kloos (2018) focus on AR's role in STEM education, noting that AR enhances learning outcomes by providing students with interactive, hands-on experiences. For instance, AR allows students to visualize complex scientific concepts in three dimensions, making abstract ideas more tangible and easier to comprehend. Their research highlights that AR fosters STEM learning by promoting exploratory and experimental activities, which are essential for deepening cognitive engagement and enhancing problem-solving abilities [55].

Dunleavy et al. (2009) review AR's application in STEM education, finding that AR's immersive capabilities allow students to interact with STEM content in ways that foster critical thinking and hands-on learning. They argue that AR provides a platform for active learning that supports deeper engagement and comprehension, particularly in complex subjects like physics and chemistry [56].

Huang et al. (2021) explore AR's effectiveness beyond traditional education by examining its role in corporate training. Their study reveals that AR can enhance knowledge acquisition and long-term retention in professional training settings, indicating AR's versatility as a learning tool in both educational and corporate environments. This research suggests that AR's learnability factors, such as intuitive navigation and real-time feedback, are beneficial across various training contexts, from academia to industry [57].

2.2.5 Learnability in Software Applications

Learnability, defined as the ease with which new users can achieve proficiency, is crucial for user satisfaction and retention in software applications. Grossman (2009) pioneered guidelines for improving learnability, demonstrating that intuitive design reduces frustration and supports faster learning. His work remains influential in software design, emphasizing that applications with high learnability foster enhanced user satisfaction and encourage sustained user engagement [58].

Blandford and Green (2008) expand on Grossman's insights, arguing that software with good learnability not only facilitates efficient task execution but also enhances user experience over time. They emphasize that design features that simplify learning positively influence user efficiency, which is particularly valuable in professional and educational software applications [59].

In AR, learnability presents unique challenges due to the complex interactions required to navigate digital elements within real-world environments. Hafsa's (2021) Learnability Enhancement Model (LEM) provides an AR-specific framework that addresses factors such as user guidance and feedback mechanisms to enhance user learning. Her model is particularly relevant for designers working with immersive technologies, offering structured strategies for improving learnability in AR applications [60].

Kron and Levin (2016) focus on evaluation methods for learnability in AR, examining how users interact with AR interfaces and proposing metrics to assess learning efficiency. They stress that evaluating AR learnability requires a nuanced approach that accounts for the spatial and cognitive challenges specific to AR. This approach is crucial for developing AR applications that are accessible and user-friendly [61].

2.2.6 Factors of Learnability in AR Applications

Learnability in AR is influenced by various factors, including mental effort, responsive feedback, and ease of interface navigation. Hafsa's (2021) Learnability Enhancement Model identifies key strategies, such as interactive tutorials and contextual cues, that help users quickly adapt to AR interfaces. Her work is foundational for improving learnability in AR by creating user-centered applications that provide seamless guidance and user-friendly interaction mechanisms [60].

Cheng and Tsai (2013) highlight the role of AR in science education, demonstrating that AR's interactive elements improve conceptual understanding by making complex topics more accessible. Their research supports the idea that AR's interactive visuals and dynamic simulations improve learning outcomes by providing students with an engaging, immersive learning experience that facilitates quick comprehension and retention [62].

2.2.7. Discussion

The literature reviewed highlights the crucial role that user experience (UX) standards, user-centered design (UCD), and learnability play in the effective development of augmented reality (AR) applications. Prior research emphasizes that while AR introduces novel possibilities for interactive and immersive experiences, the foundation of these applications must be grounded in well-established UX principles to maximize their effectiveness and ensure user satisfaction.

Various studies underscore the importance of adhering to UX standards in AR design. Panagiotis D. (2011) identifies key UX aspects such as presence, ergonomics, and overall usability, which are essential for AR systems to be both functional and user-friendly. This focus on usability is further reinforced by Jerald (2015) and the ISO/IEC 25010:2011 framework, which stress that usability, efficiency, and satisfaction are critical components in ensuring that AR systems meet user expectations. The broader research agenda set by Hassenzahl and Tractinsky (2006) and Nielsen's (1993) user-centered design methodologies also echo these sentiments, emphasizing the necessity of grounding AR experiences in established UX principles. Collectively, the literature suggests that design principles must be prioritized to ensure that AR applications are accessible and intuitive, enabling users to fully engage with the technology.

In product design, AR has been identified as a tool that can significantly enhance creativity, efficiency, and user engagement. Schumann (2020) highlights the effectiveness of integrating design models, such as the Double Diamond, into AR-supported product design workflows. This systematic approach, along with the frameworks proposed by Gao et al. (2018) and Zhang and Wang (2019), offers structured methods for incorporating AR into the design process, ensuring that AR technology is used to its fullest potential. These studies highlight the role AR plays in facilitating real-time visualization and prototyping, thus enabling faster iterations and improving usability outcomes. However, the integration of AR in product design also presents challenges, such as ensuring seamless data integration and managing the complexities of user interfaces.

User-centered design (UCD) is also pivotal in AR systems, as discussed by Duenser and Hollerer (2007), who emphasize the need for AR-specific design guidelines. The unique challenges posed by AR environments—such as interacting with both physical and digital elements—demand a user-centered approach to ensure that AR systems are both efficient and user-friendly. Henderson and Feiner (2011) further highlight the necessity of interdisciplinary collaboration in AR design, reinforcing the idea that understanding user contexts and needs is critical for creating successful AR applications. Brusilovsky and Millán (2007) also contribute to this dialogue by emphasizing the importance of integrating user insights throughout the development process, which ensures that AR systems are not only technologically advanced but also tailored to practical user experiences.

In educational contexts, AR is recognized for its potential to revolutionize learning experiences. Mena-Vargas (2019) demonstrates that AR can increase academic interest and improve educational outcomes by creating immersive, interactive environments that promote engaged participation. Ibáñez and Delgado-Kloos (2018) extend this discussion, highlighting the benefits of AR in STEM education, where it enhances the comprehension of intricate concepts by making them more accessible and visually compelling. These findings are further supported by Broll and Grimm (2015) and Tao and Wu (2018), who emphasize that AR can modernize educational practices, making learning more immersive and stimulating through the use of hands-on simulations and real-time feedback.

Learnability, a key factor in the user acceptance of software applications, is especially critical in AR environments. Grossman and Fitzmaurice (2009) laid the groundwork for understanding learnability in software design, outlining the challenges users face when first interacting with new systems. Hafsa (2021) builds on this by introducing the Learnability Enhancement Model (LEM), which identifies specific factors that influence learnability in AR, such as interface design complexity, onboarding support, and feedback clarity. Hafsa's LEM model offers strategies to reduce cognitive load and improve user performance, making it a valuable framework for AR developers. The importance of evaluating learnability is further emphasized by Cao and Zhao (2015), who propose metrics for measuring how effectively users can become proficient with new systems. Similarly, Schnabel and Brucks (2017) explore strategies for improving user interaction and learning in AR, particularly in navigation systems, where users must engage with both physical and digital elements simultaneously.

Research also highlights the impact AR has on learnability in educational contexts. Hafsa's (2021) LEM model provides a structured approach for addressing learnability in AR environments, while Wu et al. (2013) and Hsu and Ching (2013) explore how AR can improve learning efficiency and motivation. These studies suggest that AR's immersive and interactive nature can enhance learning outcomes, but they also stress the importance of thoughtful design and thorough evaluation to ensure that users can seamlessly learn and engage with AR systems. This duality—AR's potential to both enhance and challenge traditional learning processes—underscores the need for continuous research to refine AR systems for optimal learnability.

The added value of our study lies in its integration of these existing frameworks—UX standards, UCD principles, and the Learnability Enhancement Model (LEM)—into a comprehensive framework for evaluating and improving AR user experience. Our study aims to contribute to the growing body of research by proposing automated measures of learnability in AR environments, which can complement existing user-based testing methods. By emphasizing the automation of learnability assessment, our framework seeks to streamline the process of assessing user performance, learning patterns, and user feedback in real-time, providing developers with actionable insights for improving AR systems. This approach offers a more efficient and scalable solution to optimizing AR applications across various domains, from education to product design. Our contribution thus extends the existing literature by not only reinforcing the importance of UX standards and learnability in AR design but also proposing practical methods to automate and enhance the user experience in real-world settings.

Chapter 3:

Methodology

The methodology of this study, *Framework for Augmented Reality User Experience: Automated Measures of Learnability and Associated Factors*, is structured to systematically analyze and evaluate key user experience (UX) measures associated with learnability in augmented reality (AR) applications. The research involves both automated and manual data collection through user testing, automated tools, and real-time data analysis to build a robust framework for assessing AR usability. The goal is to develop a comprehensive, automated tool to enhance UX and evaluate learnability, adhering to established international standards such as ISO/IEC 25010:2011 and ISO 9241:11 (2015). The process is designed as follows:

3.1 Study Design

The study begins by analyzing key UX measures in the context of AR application design. Following this, the study develops a framework for evaluating learnability through automated tools designed with the .NET framework and Unity Platform. This methodology enables the measurement of learnability and other UX-associated factors while users interact with AR applications. Data is collected through automated extraction of user behavior and reactions, complemented by post-immersion assessments, and questionnaires to validate and refine the results.

The study concentrated on immersive experiences within educational, architectural, and tourism-based AR applications, developing an automated tool to extract key parameters associated with metrics such as learnability, effectiveness, efficiency, memorability, and user satisfaction.

The method was carried out in two phases. The first phase, referred to as "Initial Task Performance," involved participants completing a series of tasks within the AR applications. These tasks were specifically designed to evaluate the participants' ability to navigate the system, interact with virtual objects, and accomplish predefined objectives.

3.2 User Experience Measures

A central component of this study is the design and implementation of a tool that allows developers to evaluate the learnability of AR products based on international UX and HCI guidelines (ISO 9126, ISO/IEC 25010:2011). The following UX measures are targeted:

- **Learnability**: Measures how easily users can learn and complete basic tasks within an AR system, particularly on their first use.
- **Efficiency and Effectiveness**: Reflects the degree to which users achieve desired outcomes with optimal resource investment (time, cognitive effort, etc.).
- **Memorability**: Measures the ease with which users can remember how to use the system after periods of non-use. It evaluates whether users can return to the system and perform tasks without relearning.
- **Discretionary Usage**: Assesses how users engage with additional, non-core features that enhance their overall experience. It involves evaluating how frequently users interact with advanced functions beyond the primary use case.
- **Satisfaction**: Evaluates users' subjective responses and feelings about the AR application, encompassing comfort, enjoyment, and overall likeability.
- **Attractiveness**: Assesses the visual and aesthetic appeal of the AR application, focusing on how design and interface elements contribute to the overall user experience.

Table (A- 4.1): User Experience Testing Measurement factors.

UX measures	Related factors	Factor Description	Cut-Off
Learnability in User Experience (UX)	Learnability	Learnability = $\frac{E_{\{start\}} - E\_\{end\}}{T} + F$ Where: L = Learnability score E_{start} = Number of errors or unsuccessful attempts made at the start of interaction E_{end} = Number of errors or unsuccessful attempts made after a period of interaction T = Time taken to reach proficiency (or perform a task successfully without errors) F = Additional factors that affect learnability (e.g., user familiarity with similar interfaces, complexity of the task, feedback quality)	**L ≥ 0.5**: High learnability (good reduction in errors with minimal time and frustration). **L 0.1 - 0.5**: Moderate learnability (acceptable for complex systems). **L < 0.1**: Low learnability (indicating poor usability and significant room for improvement).

Table (B- 4.1): User Experience Testing Measurement factors.

UX measures	Related factors	Factor Description	Cut-Off
Effectiveness	Task Completion Rate	Effectiveness = $\frac{number\ of\ tasks\ completed\ successfully}{Total\ number\ of\ tasks\ undertaken}$ Effectiveness can be calculated by measuring the completion rate. Referred to as the fundamental usability metric, the completion rate is calculated by assigning a binary value of '1' if the test participant manages to complete a task and '0' if he/she does not.	0 or 1
	Error rate	Error Rate = $\frac{number\ of\ measure\ fail}{Total\ number\ of\ tasks\ undertaken}\ x\ 100\%$	Errors per task is 0.7, with 2 out of every 3 users making an error.
	Accuracy	Accuracy = $\frac{Number\ of\ Correct\ Actions}{Total\ Number\ of\ Actions} x100\%$ Accuracy is the proportion of correct actions taken by a user out of the total actions taken. It measures how accurately a user performs tasks, regardless of whether they complete the task successfully.	100%
	Overall Task Success Rate (OTSR)	The Overall Task Success Rate (OTSR) is a composite metric designed to provide a comprehensive view of task performance by integrating traditional usability measures. It is inspired by established methods in user experience research (Tullis & Albert, 2013; Sauro & Lewis, 2016; Nielsen, 1993; Rubin & Chisnell, 2008). The OTSR is calculated as follows: OTSR=$\frac{Task\ Completion\ Rate}{TAverage\ Completion\ Time}\ x\ (1 - Error\ score)\ x\ User\ STF\ score$	"Excellent," "Good," "Average," and "Needs Improvement."

Table (C- 4.1): User Experience Testing Measurement factors.

UX measures	Related factors	Factor Description	Cut-Off
Efficiency	Time Based Efficiency	*Time Based Efficiency*= $\frac{\sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}}}{NR}$ *Overall relative Efficiency* = $\frac{\sum_{j=1}^{R} \sum_{i=1}^{N} \frac{n_{ij}}{t_{ij}}}{\sum_{j=1}^{R} \sum_{i=1}^{N} t_{ij}} \times$ 100% N = The total number of tasks (goals) R = The number of users nij = The result of task i by user j; if the user successfully completes the task, then Nij = 1, if not, then Nij = 0 tij = The time spent by user j to complete task i. If the task is not successfully completed, then time is measured till the moment the user quits the task	100%
	Overall relative Efficiency		
Memorability	Memorability (M)	Memorability (M) = $\frac{Task\ Retention\ Score}{Time\ to\ Relearn + Error\ Rate}$	
	Task Retention Score (TRS):	Task Retention Score (TRS): = $\frac{Number\ of\ Tasks\ Correctly\ Completed\ After\ Period\ of\ Non}{Total\ Number\ of\ Tasks}$ A measure of how well users remember how to perform tasks. This could be quantified by having users' complete tasks they've learned previously and scoring them based on their success and speed. i. High Memorability: > 80% of tasks correctly recalled and completed. ii. Moderate Memorability: 50% - 80% of tasks correctly recalled. iii. Low Memorability: < 50% of tasks correctly recalled.	HM: > 80% MM: 50% - 80% LM: < 50%
	Time to Relearn (TTR)	Time to Relearn (TTR): The time it takes for users to return to their previous performance level after a period of non-use. i. High Memorability: < 5 minutes to relearn previous tasks. ii. Moderate Memorability: 5 - 15 minutes to relearn. iii. Low Memorability: > 15 minutes to relearn.	HM: < 5 minutes MM: 5 - 15 minutes LM: > 15 minutes
	Error Rate (ER)	Error Rate (ER): The number of errors made during task performance after the period of non-use. i. High Memorability: < 10% error rate upon returning to the system. ii. Moderate Memorability: 10% - 20% error rate. iii. Low Memorability: > 20% error rate.	HM: < 10% MM: 10% - 20% LM: > 20%

Table (D- 4.1): User Experience Testing Measurement factors.

UX measures	Related factors	Factor Description	Cut-Off
Discretionary usage		X = A/B A= number of **times** that specific applications are used B = number of **times they are intended to be used**	Observation of usage 100%
Satisfaction scale	Satisfaction questionnaire	System Usability Scale (SUS): A widely used 10-item questionnaire that provides a quick assessment of the usability of a system, including AR/VR applications. $\sum (A_i)/n$ A_i) = response to a question n = number of responses	SUS Scale User test
Attractiveness	Attractive interaction	QUIS Questionnaire to assess the attractiveness of the interface to users, taking account of attributes such as color and graphical design. NOTES: Issues that potentially contribute to attractiveness include: 1. Alignment of items (vertical and Horizontal), Grouping 2. Use of colors 3. Appropriate and reasonable-sized graphics 4. Use of whitespace/separators/borders 5. Animation 6. Typography 7. 3D interface	User test

3.2.1 Learnability Model

The formula calculates learnability based on the reduction in errors over time, the speed of learning, and additional factors. A higher learnability score indicates that the system is easy to learn and use, with minimal errors and a short learning curve:

- Learnability: This is the overall score that measures how easy it is for users to learn and use the system.
- E_start: The number of errors or unsuccessful attempts made at the beginning of interaction.
- E_end: The number of errors or unsuccessful attempts made after a period of interaction.
- T: The time taken to reach proficiency or perform a task successfully without errors.

- F: Additional factors that affect learnability, such as user familiarity with similar interfaces, task complexity, and feedback quality.
    - ❖ Learnability $=\frac{\text{E}_{\text{start}}-\text{E}_{\text{end}}}{\text{Time}}+F$

Interpretation:

- Error Reduction: The numerator, E_start - E_end, represents the reduction in errors over time. If users make fewer errors as they continue to interact with the system, this value will be positive, indicating good learnability.
- Time Efficiency: The denominator, T, measures the time it takes to learn the system. A smaller value indicates that users can learn quickly, which contributes to high learnability.
- Additional Factors: The F term accounts for external factors that can influence learnability. If users are familiar with similar systems or the tasks are relatively simple, these factors can contribute positively to learnability.

3.2.2. Effectiveness

Effectiveness in the context of usability and user experience refers to how well a system or tool allows users to achieve their goals. It is a measure of the accuracy and completeness with which users achieve their objectives. Here are key aspects of effectiveness:

Key Aspects of Effectiveness

- Task Success Rate: The percentage of tasks completed successfully by users. Higher success rates indicate greater effectiveness.
- Error Rate: The number of errors users make while completing tasks. Fewer errors suggest higher effectiveness.
- Accuracy: The degree to which users can perform tasks correctly without mistakes. Higher accuracy reflects greater effectiveness.
- Task Completeness: Whether users are able to complete the entire task as intended. Partial completion may indicate issues with effectiveness.
    - ❖ **Overall Task Success Rate (OTSR):**

The Overall Task Success Rate (OTSR) is a composite metric designed to provide a comprehensive view of task performance by integrating traditional usability measures. It is inspired by established methods in user experience research (Tullis & Albert, 2013; Sauro & Lewis, 2016; Nielsen, 1993; Rubin & Chisnell, 2008). The OTSR is calculated as follows:

OTSR$=\frac{Task\ Completion\ Rate}{TAverage\ Completion\ Time}\ x\ (1-Error\ score)\ x\ User\ Satisfaction\ (STF)\ Score$

Metrics and Data Collection:

- Task Completion Rate: The percentage of tasks successfully completed by users within the given time frame.
- Average Completion Time: The average time taken to complete tasks.
- Error Rate: The number of errors or mistakes made during task performance.
- User Satisfaction: Subjective ratings of the experience, often gathered through post-task questionnaires.

Compute the Difference or Percentage Change:

Absolute Difference:

Difference = $OTSR_{Session\ 2}$- $OTSR_{Session\ 1}$

A positive value indicates improvement, while a negative value suggests a decline in performance.

Percentage Change:

$$\text{Percentage Change} = \frac{OTSR_{Session\ 2} - OTSR_{Session\ 1}}{OTSR_{Session\ 1}} x\ 100\%$$

This formula shows how much the OTSR has increased or decreased as a percentage of the first session's score.

3.2.3. Efficiency

Efficiency in the context of usability and user experience measures how quickly and resourcefully users can achieve their goals using a system or tool. It is concerned with the amount of time, effort, and resources required to complete tasks. High efficiency means that users can accomplish their tasks with minimal time and effort.

Key Aspects of Efficiency

- Time to Completion: The amount of time it takes for users to complete a task from start to finish. Shorter times generally indicate higher efficiency.
- Number of Actions: The total number of actions or steps required to complete a task. Fewer actions suggest greater efficiency.

3.2.4. Memorability

The aim of this study was to assess usability measures, with a particular focus on the importance of memorability in augmented reality (AR) applications. It explored how easily users could recall and interact efficiently with the system after a period of non-use. Memorability is a crucial component of usability, especially for systems that are not used on a daily basis or are part of intermittent workflows. Systems with high memorability enable users to quickly re-engage, reducing the need for retraining, lowering cognitive load, and contributing to an overall positive user experience.

The memorability was measured using a sample of 20 participants across different age groups and sectors, including education, tourism, and architecture. The AR applications used in the study were designed for specific tasks within these domains, such as exploring virtual learning environments, interacting with historical landmarks, and navigating architectural models.

After a one-month interval without using the application, participants returned for the second phase, a "Follow-Up Session," where they completed the same tasks. During this follow-up session, key metrics related to memorability were recorded, including task retention, time to relearn, and error rates.

To assess memorability, several key measures were utilized:

Task Retention measured participants' ability to recall and complete tasks accurately after a period of non-use. This metric was calculated as the percentage of tasks completed correctly during the follow-up session compared to the initial session. The Task Retention Score (TRS) was determined using the formula:

$$TRS = \left(\frac{\text{Number of Tasks Correctly Completed After Non-Use}}{\text{Total Number of Tasks}}\right) \times 100$$

Memorability was classified based on retention levels: a score above 80% indicated high memorability, between 50% and 80% indicated moderate memorability, and below 50% indicated low memorability.

Time to Relearn (TTR) measured the time required for participants to regain their initial level of proficiency after the period of non-use, with the aim of determining how quickly users could recall task steps in the system. The formula for TTR was:

$$\text{TTR}=\text{Time Taken to Relearn Tasks to Initial Proficiency}$$

Memorability was classified by the time to relearn: less than 5 minutes indicated high memorability, 5 to 15 minutes indicated moderate memorability, and over 15 minutes indicated low memorability.

Error Rate After Non-Use was recorded to assess the number of mistakes participants made when completing tasks in the follow-up session. This metric provided insights into the accuracy of task recall and execution. The Error Rate (ER) was calculated as:

$$ER = \left(\frac{\text{Number of Errors}}{\text{Total Number of Actions}}\right) \times 100$$

A lower error rate corresponded with higher memorability. A rate below 10% indicated high memorability, between 10% and 20% indicated moderate memorability, and above 20% indicated low memorability.

Cognitive Load was measured through a follow-up questionnaire, where participants rated the mental effort required to recall and use the AR system after a break. The cognitive load scale ranged from 1 to 5, with 1 representing minimal effort and 5 indicating high cognitive demand.

An **Overall Memorability Score** was then calculated by combining Task Retention Score (TRS), Time to Relearn (TTR), and Error Rate (ER). This composite score provided a holistic view of each group's ability to remember and efficiently use the system after non-use. The formula used was:

$$\text{Overall Memorability Score} = \frac{TRS + (1 - \frac{TTR}{\text{Max TTR}}) + (1 - \frac{ER}{\text{Max ER}})}{3}$$

The results revealed different levels of memorability among groups: Group A achieved an overall score of 0.65, indicating moderate memorability; Group B achieved 0.81, indicating good memorability; and Group C achieved 0.91, reflecting high memorability.

Overall Memorability Score

To summarize the results, an **Overall Memorability Score** was calculated by combining the Task Retention Score (TRS), Time to Relearn (TTR), and Error Rate (ER). This provided a comprehensive understanding of each group's ability to remember and efficiently use the system after a period of non-use.

Overall Memorability Score=TRS+(1−TTR/Max TTR)+(1−ER/Max ER)/3

- **Group A**: 0.65 (Moderate Memorability)
- **Group B**: 0.81 (Good Memorability)
- **Group C**: 0.91 (High Memorability

3.2.5. Satisfaction

User Satisfaction (Subjective Measures): User satisfaction was evaluated using post-experience questionnaires or interviews, where participants rated their satisfaction with the AR/VR experience. These questions covered various aspects, including ease of use, comfort, and overall enjoyment. A key tool used for measuring satisfaction was the System Usability Scale (SUS), a widely recognized 10-item questionnaire that helps gauge users' perceptions of the system's usability.

Each question within the SUS is designed to evaluate different aspects of the user experience, such as the overall usability of the application, the accuracy of the extended reality features, and the effectiveness of the training content. The SUS uses a five-point Likert scale for responses, ranging from "Strongly agree" to "Strongly disagree."

The System Usability Scale (SUS) is a widely used tool for assessing the usability of a system or product. It includes a 10-item questionnaire that users complete after interacting with the system. Each item is rated on a 5-point scale, with scores adjusted based on whether the item is positively or negatively worded. The final score is calculated by multiplying the total score by 2.5, converting it to a scale from 0 to 100. Higher scores indicate better usability, providing a clear and effective measure of user satisfaction and system performance.

By administering the SUS to 20 users and calculating the average score, we were able to gain a clear understanding of the overall usability of the AR app.

3.2.6. Attractiveness

This section of the study focused on evaluating the attractiveness of AR applications across various domains, including education, tourism, and architecture. The goal was to quantify how the design elements of AR systems influence user perceptions, particularly in terms of usability, emotional response, and overall satisfaction. The study of attractiveness was integrated into the overall UX framework and used a combination of **automated tools** and **user feedback** to gather data. The following methodologies were applied to ensure a comprehensive evaluation:

1. **User Surveys and Questionnaires**: The **Questionnaire for User Interaction Satisfaction (QUIS)** was administered to participants. This 27-item questionnaire was created and designed to measure subjective satisfaction with various aspects of the AR interface, including its visual appeal, terminology, system information, learning, and system capabilities. The questionnaire was divided into five sections (Overall Satisfaction, GUI, Terminology and System Information, Learning, System Capabilities), allowing for a granular analysis of each design component.
2. **User Engagement Metrics**: Through the integration of the **Unity Platform** and **Firebase Realtime Database**, automated tools tracked user interactions, including time spent on the application, navigation patterns, and bounce rates. These metrics provided objective data to complement subjective feedback, offering insights into how the attractiveness of the design impacted user engagement and retention.
3. **Discretionary Usage**: Discretionary Usage refers to the extent to which users interact with optional features within a system, going beyond its core functions. It reflects users' willingness and interest in exploring additional capabilities that enhance their overall experience. Measuring discretionary usage provides valuable insights into the system's ability to foster deeper engagement and increase user satisfaction by offering features that encourage exploration and interaction.
   In this framework, **discretionary usage** is integrated alongside other critical usability metrics—such as **effectiveness**, **efficiency**, **memorability**, **satisfaction**, and **attractiveness**—to create a comprehensive evaluation of user experience in AR applications. By capturing both core interactions and how users engage with optional features, this data-driven approach ensures that AR systems are not only functional but also engaging and satisfying, ultimately designed to meet and exceed user expectations.

The following key aspects of attractiveness were evaluated:

1. **Visual Design**: This included the overall aesthetic quality of the AR system, such as color schemes, typography, imagery, and layout. Users were asked to evaluate how well the visual design aligned with the purpose and context of the application, as well as its influence on ease of navigation and interaction.
2. **Consistency**: The study examined the consistency of design elements across the interface, ensuring that users experienced a cohesive and professional design throughout. Participants provided feedback on whether they encountered disjointed design elements or if the interface maintained a uniform visual style across different sections and features.
3. **Usability Integration**: Attractiveness was assessed not only for its visual appeal but also for how it supported or detracted from usability. The goal was to determine whether the aesthetic elements enhanced the user's ability to interact with the system or, conversely, if they introduced confusion or complexity. Feedback was gathered on how well the interface balanced aesthetics with functionality.
4. **Emotional Impact**: The ability of the design to evoke positive emotional responses was a significant factor in the study. Users were asked to reflect on how the visual design made

them feel—whether it was engaging, motivating, and enjoyable, or if it left them feeling indifferent or frustrated.

3.3. Framework Development

3.3.1. Augmented Reality User Experience Framework Development

A Learning testing framework was developed that used Python-Django for testing AR application performance. The framework was developed based on HTML, CSS, and Python programming language using Django framework. The platform database used Firebase databases. Figure 3.1 shows the general architecture of the Learnability framework.

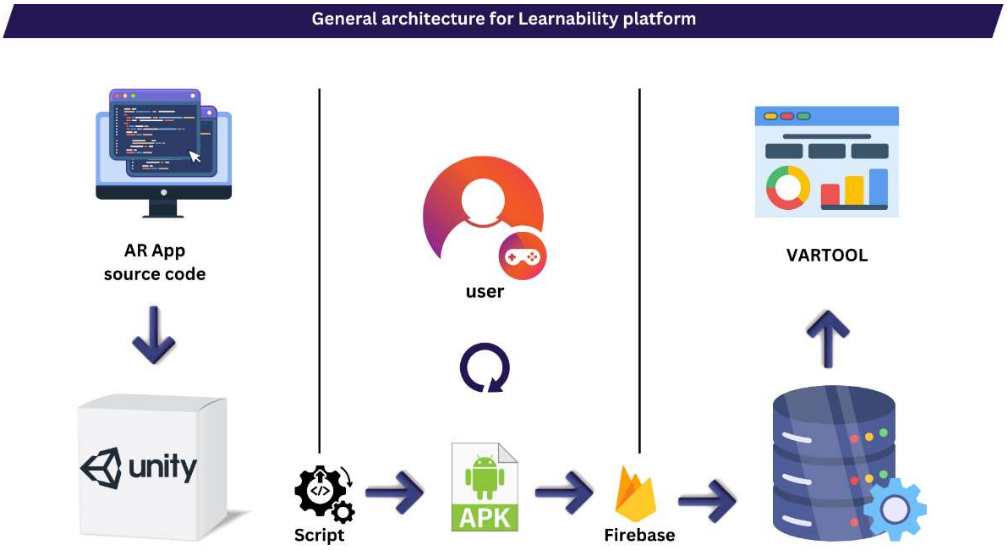

Figure 3.1: General architecture for Learnability platform

The framework composed of three main components: The platform development layer based on Html, CSS and Python programing language; The Unity platform and the Firebase database layer, in which the Learnability measures variables were stored and analyzed.

3.3.2. VARTOOL Tool Design

The flow architecture is illustrated in Figure 3.1. The design is composed of three main stages: Two stages for accommodation Unity and firebase to collect data, and one stage for data analysis and visualization. By inserting packages into the source code, the apps will be modified to ensure compatibility with Firebase, enabling the extraction of user experience measures variables such as task completion, execution time, device details, and user information.

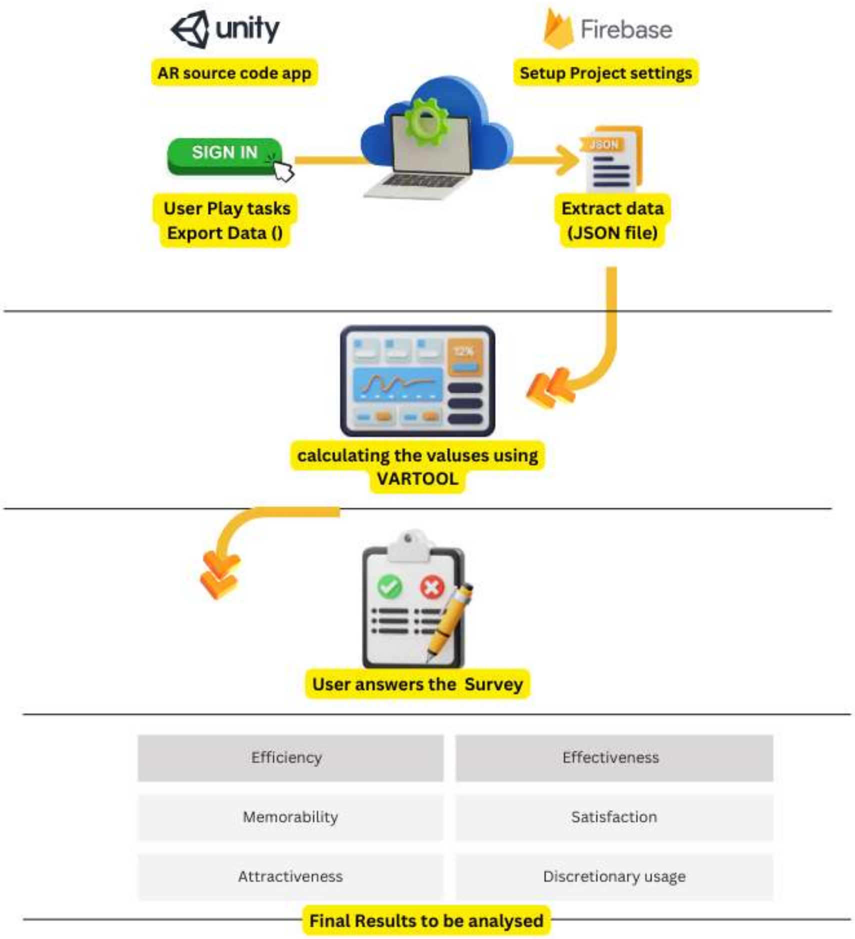

Figure 3.2: AR UX Farmwork flow architecture

3.3.3. Data Analysis

The data analysis VARTOOL tool was designed to provide a user experience analysis for the AR applications by conducting comparative and statistical analysis of user experience measures parameters. The VARTOOL connects to the framework database to access the collect learnability variables data and runs the predefined analysis functions. The analysis functions were developed in reference to Nielson usability standards and guidelines of ISO/9126, ISO 9241:11 (2015), and ISO/IEC 25010:2011. In this research, the applications were assessed in terms of task completion, execution time, and overall task completion rate. The VARTOOL analysis workflow is illustrated in Figure 3.2.

The tool flow architecture, consists of the following steps: first, the application's source code is imported into Unity, where C# scripts are added for measuring and storing performance metrics and for transmitting these metrics to Firebase. The application is then exported as an APK for end-user deployment. The tool interacts with Firebase as described in section 3.3.2, after which a questionnaire is presented to the user. The questionnaire results are then analyzed and compared with the learnability factors.

3.3.4. Experimental Setup

The experiment was conducted in two different settings to evaluate the usability of the three AR applications—EasyApp, Farah, and Dar Al Consul (DAC). For EasyApp, the experiment was open-ended, allowing participants to interact with the app in a more flexible environment. Participants were given the freedom to explore the app's architecture features on their own, with minimal guidance, in a controlled, distraction-free room. This approach aimed to observe the natural flow of interaction and the ease with which users could adapt to the app's functionalities. In contrast, the experiments for Farah App and DAC followed a closed setup. For Farah, the users were assigned specific tasks—coloring a picture, finding hidden stars, and watching an educational video. These tasks were designed to guide the users through the app's key functionalities, and their progress was monitored closely. Similarly, DAC, which focuses on tourism, presented a more structured set of interactions to the participants. Both Farah and DAC experiments were carefully controlled, with clear instructions provided to users before starting the tasks. Data collected from all three environments were recorded in real-time, allowing for detailed analysis of user behavior and performance.

3.4. Framework Testing and Evaluation

3.4.1. Study Sample

The user experience framework was tested on a sample of 20 user playing 3 locally AR applications including fields of education, tourism and architecture The user experience framework was tested on a sample of 20 users playing 3 local AR applications including fields of education, tourism and architecture. For each application, three tasks were identified and evaluated: in the Farah App (Coloring, Finding Stars, Watching Videos), in DAC (Getting a Gift from Santa Claus, Rotating Marbles, Displaying Information), and in Easy AR (Resizing object, rotating object, deleting object). A total of 60 attempts were conducted, each tested twice over one month.
AR Applications

- Farah App

The study used three different augmented reality applications, including Farah shown in Figure (1), an interactive app aimed at raising awareness about children's rights in Palestine. The app is linked to a coloring book and is used by children to learn about their rights through activities such as coloring, finding hidden stars, and watching educational videos. Farah and Adam are the main characters in the app, siblings from Jerusalem who talk daily about their future dreams and rights. The app is designed to evaluate ease of learning and cognitive processing in children and we aim to study the performance metric of this apps.

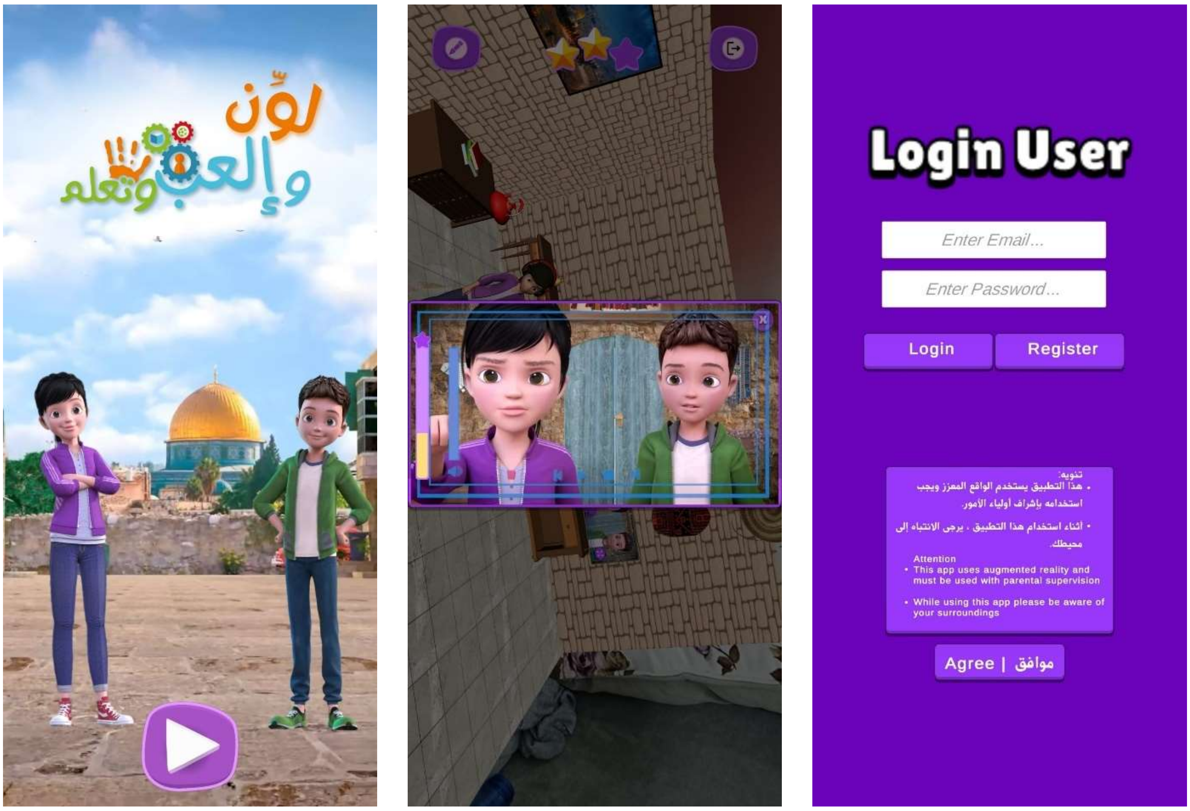

Figure 3.3: Farah App

- Dar-Al Consul Jerusalem

An application that brings the sights and sounds of Jerusalem's old city, through an EU, UN-Habitat, CTS and Al-Quds University project at/in the center of the old city. The aim of the project is bringing history to life to enrich the journey of the visitors through Dar Al Consul Community and Civic Center's discovered and reconstructed rooms and spaces. Visitors are now able to get a glimpse of the past and the rich history of Dar Al Consul through different ages and eras. The AR experience covers the water canal which once streamed along the Cardo of the old city of Jerusalem.

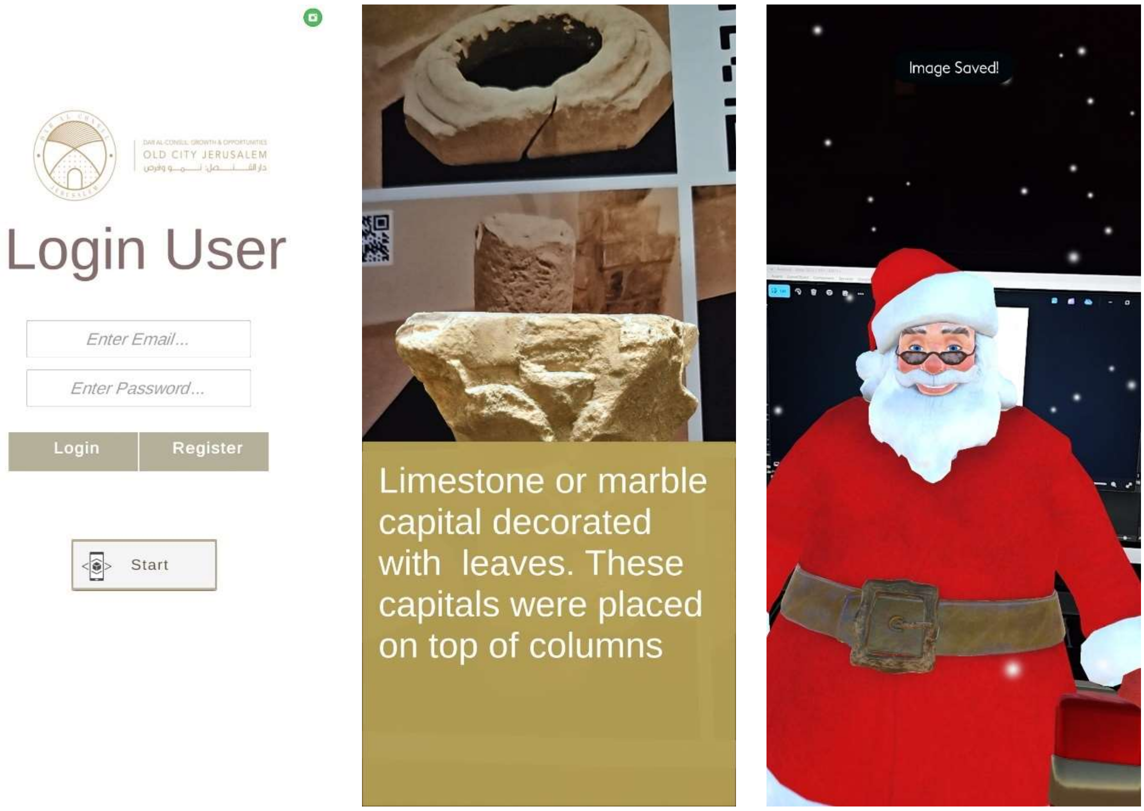

Figure 3.4: Dar-Al Consul App

- Easy AR (Architecture)

Easy AR is a Unity package designed to allow Augmented Reality applications to be easily set up by users without requiring any coding skills. The features of Easy AR have been improved with AR Foundation 4+ versions, and 8 new utility scenes have been brought, including AR Single Vertical & Horizontal Placement, Multiple Object Placement, AR Measurement, AR Tiling, AR Image Target, AR Face Filter, and AR Portal. Easy AR allows developers to implement AR features in applications without advanced programming skills. It is designed for ease of use, making it accessible to beginners while providing robust tools for advanced users.

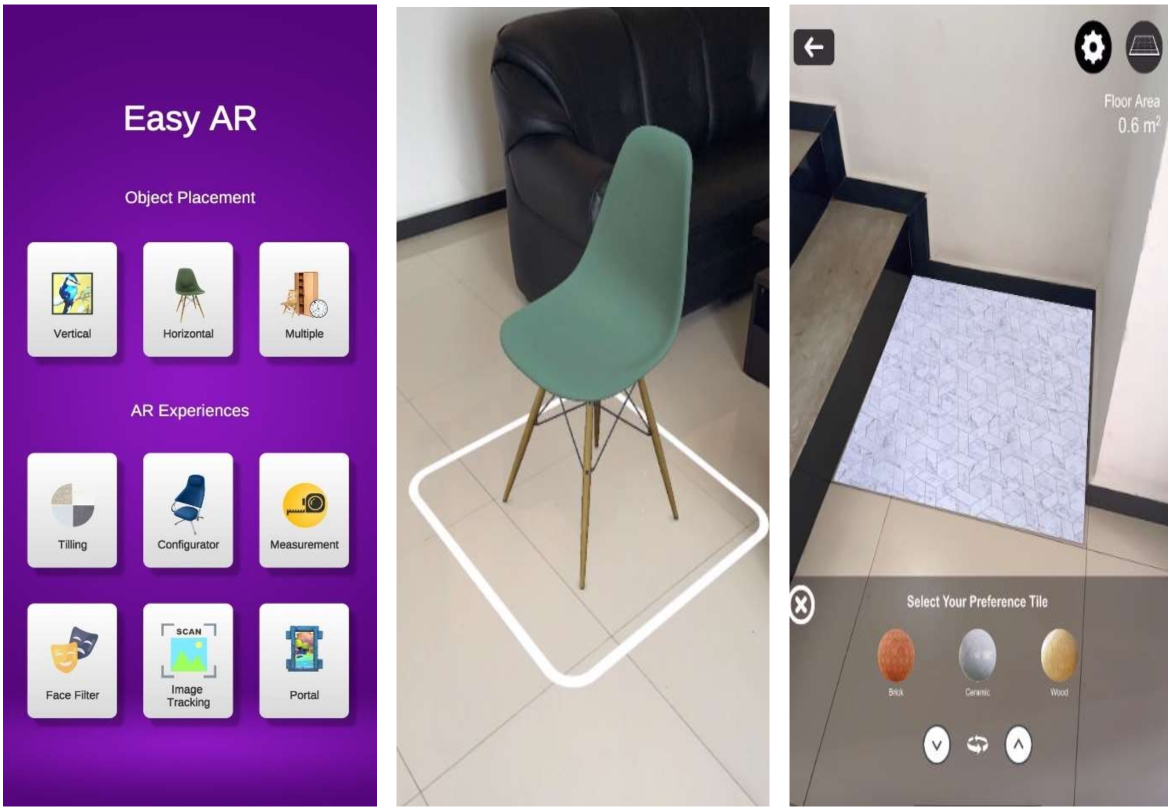

Figure 3.5: Easy AR App

For institutions' privacy protection and ethical considerations, the selected users were coded into anonymous codes. Table 3.4 shows the coding distribution by business type.

Table 4.2 Users categorized by Age

Age Category	Code	Tested apps
Children (6-12 years)	C1-C10	Farah, Dar, Easy AR
Teenagers (13-18 years)	T1, T2, T3,	Farah, Dar, Easy AR
Adults (19-64 years)	A1, A2, A3, A4, A5, A6	Farah, Dar, Easy AR

3.4.2. Data collection

- Learnability Dashboard

Data collection was fully automated and stored in databases using Unity-Firebase integration. The Tool-Firebase integration then facilitated the display of this data on the tool's dashboard. After the user completes the application, they are presented with a questionnaire, and their responses are stored in the database.

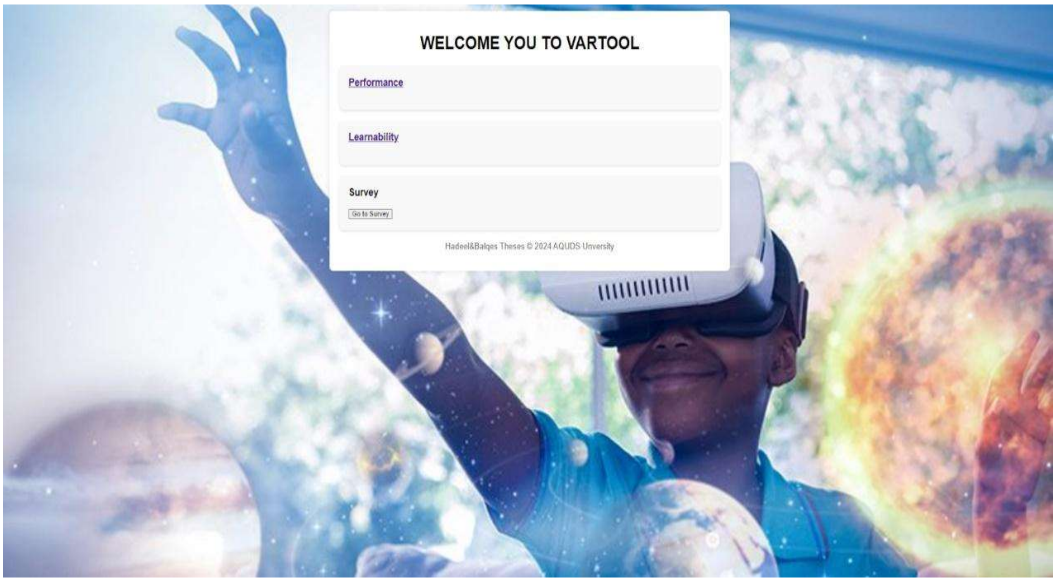

Figure 3.6: VARTOOL Main Screen

- questionnaire Page

Once the user completes the application, the tool will present them with a questionnaire. After the user fills out the SUS and QUIS questioners, their responses will be stored in the database.

3.4.3. Data Analysis and Visualization

User experience analysis was carried out using a VARTOOL tool, with results displayed through both tabular and graphical visualizations. A results presentation dashboard was integrated into the framework and linked to a real-time database. Descriptive statistical analysis was employed to present general statistics, including averages, percentages, and frequencies.

3.4.4. VARTOOL Results Visualization

VARTOOL dashboard is illustrated in Figure 3.7, developed using Django and Python, presents a user-friendly interface for analyzing usability metrics related to user task performance. The layout is neatly organized into multi panels, each displaying key performance data for separate login sessions. Each panel provides essential details such as Total Task Time, Error Rate, Task Completion Rate, Average Time to Complete, and the Overall Success Task Rate (OSTR), etc.

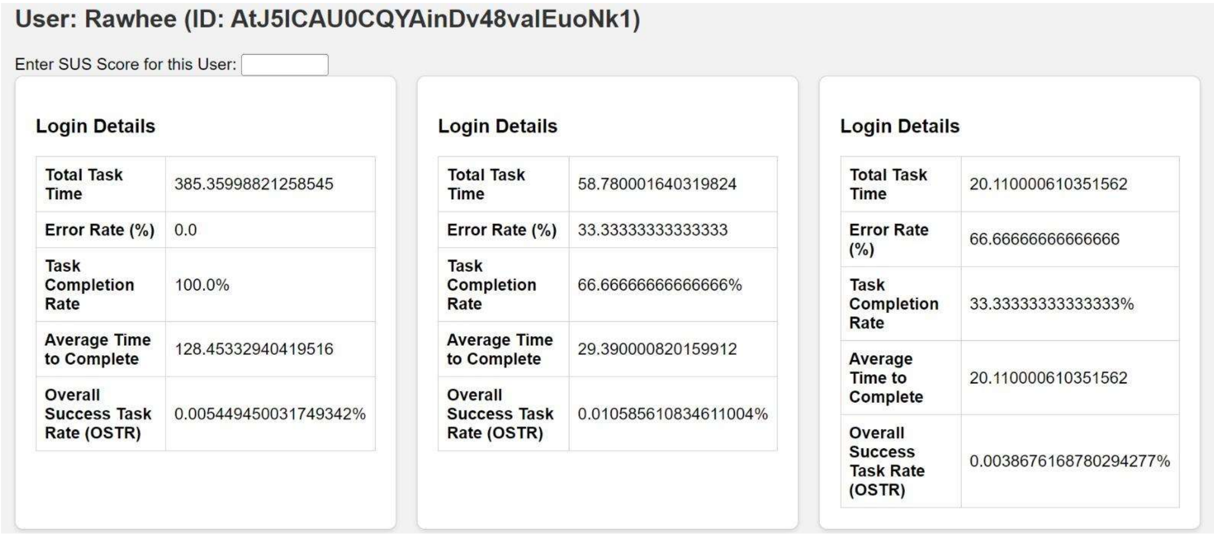

Figure 3.7: Results Dashboard

The design leverages Django's capabilities for dynamic data handling and displays calculations in a clean, tabular format for easy comparison across sessions. The numerical data offers a comprehensive view of user performance, enabling detailed analysis and insights into task efficiency, error frequency, and overall user success. This straightforward layout enhances usability and ensures clear visibility of critical metrics for informed decision-making.

Chapter 4:

Results and Analysis

In this chapter, the results and analysis of the usability testing for three distinct AR applications-Farah App (Educational), Dar Al Consul (Tourism), and EasyApp (Architecture)-are presented. These applications were evaluated using the VARTOOL framework, which focuses on assessing the learnability, effectiveness, and user engagement in augmented reality interfaces. The study involved 20 volunteer users, spanning a range of age groups, who interacted with the applications on a smartphone device. Alongside automated data collection, Participant responses was gathered through a detailed questionnaire to gain deeper insights into the participants' experiences.
The analysis presented in this chapter delves into the key usability metrics—such as task completion rates, error rates, and task retention scores—while also considering the impact of age and session improvements on the overall usability of each application. The findings are discussed in relation to the learnability of the applications, with specific attention given to the ease of use, engagement levels, and users' ability to quickly adapt to the features within each AR environment. The results aim to contribute to the ongoing development of intuitive, user-friendly AR interfaces, particularly in the domains of education, tourism, and architecture.

4.1 Study Results

The VARTOOL was used to evaluate the learnability of three distinct AR applications—Educational, Tourism, and Architecture—across a diverse group of 20 volunteer users, spanning various age ranges, who interacted with these apps on a smartphone device. In addition to gathering learnability data during the evaluation, the study also collected user feedback through a detailed questionnaire. The responses provided insights into the user experience and ease of learning associated with each application.

4.2 Learnability Analysis

Farah App developed to track three tasks, we added a full package to the Farah App and identified the tasks to be measured. Users are assigned complete tasks, after user finish the game and press exit, directly the recorded data will be transferred to the real-time database on firebase and presented as a hierarchy Tree as in Figure 3.7 presented.

Table 4.3 The learnability analysis of Farah, DAC and EasyApp applications using the automated framework

User	Age	Farah App				Dar AL Consul				EasyAPP			
		E start	E end	T	Learnability	E start	E end	T	Learnability	E start	E end	T	Learnability
A1	60	0.00	0.00	8.00	0.50	0.00	0.00	1.00	0.50	0.00	0.00	1.00	0.50
A2	37	1.00	0.00	5.00	0.69	0.00	0.00	1.00	0.50	0.00	0.00	2.00	0.50
A3	33	0.00	1.00	7.00	0.36	0.00	0.00	1.00	0.50	0.00	0.00	2.00	0.50
A4	29	1.00	1.00	5.00	0.50	0.00	0.00	1.00	0.50	0.00	0.00	5.00	0.50
A5	29	0.00	1.00	6.00	0.33	1.00	0.00	1.00	1.20	0.00	0.00	2.00	0.50
A6	28	0.00	0.00	13.00	0.50	0.00	1.00	2.00	-0.15	0.00	0.00	2.00	0.50
T1	18	1.00	0.00	8.00	0.63	0.00	1.00	1.00	-0.37	0.00	0.00	1.00	0.50
T2	13	0.00	0.00	8.00	0.50	0.00	0.00	2.00	0.50	0.00	0.00	2.00	0.50
T3	13	0.00	0.00	7.00	0.50	0.00	0.00	1.00	0.50	0.00	0.00	1.00	0.50
C1	10	1.00	0.00	7.00	0.64	0.00	0.00	1.00	0.50	0.00	0.00	1.00	0.50
C2	9	0.00	0.00	8.00	0.50	0.00	0.00	2.00	0.50	0.00	0.00	2.00	0.50
C3	9	0.00	0.00	16.00	0.50	0.00	0.00	1.00	0.50	0.00	0.00	1.00	0.50
C4	8	0.00	0.00	7.00	0.50	0.00	0.00	1.00	0.50	0.00	0.00	2.00	0.50
C5	6	0.00	0.00	10.00	0.50	0.00	0.00	2.00	0.50	0.00	0.00	2.00	0.50
C6	6	1.00	1.00	4.00	0.50	0.00	0.00	5.00	0.50	0.00	0.00	2.00	0.50
C7	9	2.00	1.00	4.00	0.76	0.00	0.00	2.00	0.50	0.00	0.00	2.00	0.50
C8	8	1.00	0.00	2.00	1.00	0.00	0.00	2.00	0.50	0.00	0.00	3.00	0.50
C9	7	0.00	3.00	3.00	-0.40	0.00	0.00	2.00	0.50	0.00	1.00	3.00	0.13
C10	5	2.00	0.00	8.00	0.74	0.00	0.00	3.00	0.50	0.00	2.00	4.00	0.05

The results in Table 4.1 offer valuable insights into the learnability of the three AR applications—Farah App, Dar Al Consul (DAC), and EasyApp—by assessing user interactions, task completions, and efficiency. The learnability score, with values closer to 1, indicates better adaptability and ease of learning, while scores near 0 or negative suggest challenges in usability.

Farah App: The learnability scores for Farah App shows considerable variation. For older users, such as A1 (age 60) and A2 (age 37), the scores ranged from 0.5 to 0.7, suggesting moderate learnability. This indicates that older users may face some challenges in adapting to the app, potentially due to complex or less intuitive elements. Conversely, younger users, such as C8 (age 8), demonstrated high scores like 0.9964, which indicate ease of learning and adaptability. This supports the idea that younger users tend to have higher adaptability, possibly because they are more familiar with modern technology or have greater cognitive flexibility. However, extreme cases like C9 (age 7), who scored -0.4005, suggest specific usability issues that could cause confusion or difficulty for certain users, highlighting the need for the app to address these challenges and ensure it is intuitive for all age groups.

Dar Al Consul (DAC): The learnability scores for DAC are relatively consistent, with most users scoring around 0.5, indicating moderate ease of learning. However, there are exceptions, such as T1 (age 18) and A6 (age 28), who scored negatively (-0.3681 and -0.1450, respectively). These negative scores indicate potential difficulties in completing tasks, suggesting that DAC may require further refinement, especially in its interface or guidance features, to improve the experience for users in these age groups. The relatively uniform scores also suggest that DAC does not cater strongly to any particular age group, achieving moderate usability across the board but lacking significant strengths in helping users learn quickly or effectively.

EasyApp: The learnability scores for EasyApp are similar to DAC, with most users scoring around 0.5, and showing limited variation across age groups. Users such as A5, A6, and C10 have scores close to zero, which implies they were able to complete tasks but faced some challenges, possibly requiring more time or encountering minor obstacles. However, C9 (score: 0.1298) and C10 (score: 0.0530) had lower scores, which suggest difficulties in task retention or navigation. This points to areas within EasyApp's interface that might need improvement to help users better understand the app and navigate through tasks more efficiently.

In summary, Farah App demonstrated a wider variation in learnability, with younger users showing a much higher ease of learning compared to older users. This suggests that the app is more suitable for a younger audience, possibly due to its design, complexity, or the cognitive demands it places on users. On the other hand, DAC and EasyApp showed more consistent but moderate learnability scores across all users, suggesting they offer a more balanced usability experience that could appeal to a broader audience. However, both apps also exhibit areas where learnability can be improved—DAC in particular, with negative scores for some users, indicates that its interface may need refinement to support users in completing tasks more intuitively. Both DAC and EasyApp could benefit from addressing these usability issues, especially for users who faced difficulties or had negative learnability scores.

These findings suggest that while Farah App excels in adaptability for younger users, DAC and EasyApp demonstrate broader usability, but require additional adjustments to enhance learnability across age groups and improve the experience for all users.

4.3 Effectiveness Results Analysis

The effectiveness analysis in Table 4.2 compares three applications—Farah App, Dar Al Consul (DAC), and EasyApp—across task completion rate, error rate, and accuracy between two sessions. This analysis helps assess each application's user performance stability over time and identifies areas for potential improvement.

Task Completion Rate: Farah App showed an increase in task completion rate from 0.783 in Session 1 to 0.883 in Session 2, marking a 0.1 improvement, indicating that users became more proficient with the app over time. In contrast, DAC exhibited a slight decrease from 0.983 to 0.966, a difference of -0.017, suggesting minor challenges in sustaining initial completion rates. EasyApp also showed a slight decline in task completion rate from 1.000 to 0.950, a 0.05 reduction, though it started with a perfect completion rate in Session 1. The average task completion rate across all apps improved by 0.011, from 0.922 in Session 1 to 0.933 in Session 2, suggesting a slight overall increase in task completion proficiency with repeated use.

Error Rate: Error rates offer insights into user precision when using each application. Farah App users demonstrated an improvement, reducing their error rate from 0.216 in Session 1 to 0.116 in Session 2, with a difference of -0.1, indicating that users made fewer errors as they became more familiar with the app. However, DAC users experienced a substantial increase in error rate from 0.016 to 0.333, a difference of 0.317, which could point to potential usability challenges in DAC that became more apparent over time. EasyApp showed a minor increase in error rate, rising from 0.000 in Session 1 to 0.050 in Session 2. On average, the error rate increased from 0.077 to 0.166 across applications, with a difference of 0.089, indicating a slight increase in errors overall, mainly driven by the performance on DAC.

Accuracy: Accuracy remained constant at 1.000 across both sessions for all three applications, with no difference noted. This stability suggests that despite variations in task completion rates and error rates, users were able to maintain accuracy in task performance consistently across sessions, indicating that the apps supported a reliable level of precision in user interactions.

Furthermore, Farah App showed a positive trend in both task completion and error reduction, suggesting improved user familiarity and effectiveness over time. DAC, while maintaining high accuracy, saw a notable increase in error rate and a slight decrease in task completion rate, highlighting areas for potential enhancement to support sustained performance. EasyApp demonstrated a minor reduction in task completion rate and a slight increase in error rate between sessions, though accuracy remained unaffected. The slight increase in overall task completion rate across apps (0.011) suggests a trend of improved effectiveness with repeated use, while the increase in error rate (0.089) indicates the potential for minor usability refinements, particularly in DAC, to ensure consistent user performance across sessions.

Table 4.4 The effectiveness analysis of Farah, DAC and EasyApp applications using the automated framework

app	Task Completion Rate			Error rate			Accuracy		
	Session 1	Session 2	Difference	Session 1	Session 2	Difference	Session 1	Session 2	Difference
Farah App	0.78	0.88	0.10	0.22	0.12	-0.10	1.00	1.00	0.00
Dar AL Consul	0.98	0.97	-0.02	0.02	0.33	0.32	1.00	1.00	0.00
EasyAPP	1.00	0.95	-0.05	0.00	0.05	0.05	1.00	1.00	0.00
AVG	0.92	0.93	0.01	0.08	0.17	0.09	1.00	1.00	0.00

The analysis in Table 4.3 provides an interpretation of the Overall Task Success Rate (OTSR) for users of different ages across three applications: Farah App, Dar Al Consul (DAC), and EasyApp, over two sessions. The OTSR indicates how effectively users can complete tasks within each application, with higher rates and positive percentage changes suggesting improved task success and learnability over time.

Farah App: For Farah App, there is a noticeable improvement in OTSR for most users from Session 1 to Session 2. Notable increases are seen among younger users, such as C4 (age 8) with a 2345% increase and C5 (age 6) with an 837% increase. These significant improvements indicate that younger users found the app easier to learn and perform tasks over time. Conversely, older users displayed more modest improvements or even decreases, as seen in user C10 (age 5) with a -67% change. The average OTSR change for Farah App across all users was positive, with a percentage increase of 108%, indicating that most users improved their task success rate in the second session, suggesting overall improved familiarity with Farah App.

Dar Al Consul (DAC): The OTSR for DAC shows mixed results. While some users, such as A2 (age 37) with a 75% improvement and A3 (age 33) with a 28% increase, demonstrated gains in task success, others showed declines. Notably, younger users like C5 (age 6) and C6 (age 6) exhibited significant decreases of -38% and -60%, respectively. This suggests that while some age groups found the app increasingly accessible over time, younger users may have encountered persistent difficulties with DAC's interface or task structure. On average, DAC had a more modest improvement in OTSR, with a 264% increase, indicating varied success among different age groups and a potential need for interface modifications to support younger users.

EasyApp: EasyApp displayed a generally positive trend, with users across various age groups showing significant improvements. For instance, A2 (age 37) had a 94% increase, and A4 (age 29) had a 205% increase in OTSR between sessions. However, a few users, such as C8 (age 8) and C10 (age 5), showed declines of -92% and -44%, respectively, indicating potential usability challenges. Despite these exceptions, the majority of users displayed improvements, with an overall average percentage increase of 264% in OTSR. This suggests that EasyApp is relatively intuitive for most age groups, although adjustments may be needed to better support very young users.

Across all three applications, younger users generally demonstrated more significant improvements in OTSR between sessions, indicating strong learnability and increased proficiency with repeated use. Farah App had the highest consistency in improvement across all age groups, while DAC showed mixed results, with some users facing challenges in maintaining high success rates, especially younger participants. EasyApp demonstrated overall positive changes, though a few young users encountered difficulties, which may highlight specific usability issues. These results suggest that while each application is learnable for most users, targeted improvements, especially in DAC and EasyApp, could enhance accessibility and task success for younger users.

Table 4.5 The Overall Task Success Rate (OTSR) based on user age

User	Age	Farah App				Dar AL Consul				EasyAPP			
		OTSR–Session 1	OTSR–Session 2	Difference	Percentage Change	OTSR–Session 1	OTSR–Session 2	Difference	Percentage Change	OTSR–Session 1	OTSR–Session 2	Difference	Percentage Change
A1	60	0.0044	0.0047	0.0002	0.0500	0.0194	0.0266	0.0072	0.3700	0.0149	0.0243	0.0094	0.6300
A2	37	0.0015	0.0050	0.0035	2.2900	0.0164	0.0288	0.0124	0.7500	0.0200	0.0388	0.0188	0.9400
A3	33	0.0021	0.0035	0.0014	0.6500	0.0221	0.0283	0.0062	0.2800	0.0150	0.0327	0.0177	1.1800
A4	29	0.0030	0.0015	-0.0014	-0.4800	0.0206	0.0180	-0.0026	-0.1300	0.0107	0.0325	0.0218	2.0500
A5	29	0.0022	0.0022	0.0000	0.0000	0.0223	0.0222	0.0000	0.0000	0.0056	0.0068	0.0012	0.2200
A6	28	0.0023	0.0037	0.0014	0.6100	0.0157	0.0120	-0.0037	-0.2400	0.0174	0.0163	-0.0011	-0.0700
T1	18	0.0022	0.0022	-0.0001	-0.0200	0.0133	0.0127	-0.0005	-0.0400	0.0117	0.0341	0.0224	1.9100
T2	13	0.0020	0.0039	0.0019	0.9800	0.0193	0.0207	0.0013	0.0700	0.0287	0.0221	-0.0066	-0.2300
T3	13	0.0034	0.0034	0.0000	0.0000	0.0103	0.0317	0.0214	2.0700	0.0130	0.0255	0.0125	0.9600
C1	10	0.0041	0.0029	-0.0012	-0.2900	0.0196	0.0221	0.0025	0.1300	0.0345	0.0292	-0.0053	-0.1500
C2	9	0.0015	0.0289	0.0046	3.0300	0.0288	0.0118	-0.0170	-0.5900	0.0270	0.0289	0.0020	0.0700
C3	9	0.0041	0.0316	0.0275	6.6300	0.0196	0.0197	0.0001	0.0000	0.0129	0.0316	0.0187	1.4500
C4	8	0.0012	0.0304	0.0291	23.4500	0.0223	0.0210	-0.0013	-0.0600	0.0241	0.0304	0.0062	0.2600
C5	6	0.0030	0.0279	0.0249	8.3700	0.0231	0.0143	-0.0088	-0.3800	0.0119	0.0279	0.0160	1.3500
C6	6	0.0029	0.0149	0.0120	4.0700	0.0300	0.0119	-0.0182	-0.6000	0.0155	0.0149	-0.0006	-0.0400
C7	9	0.0019	0.0120	0.0101	5.3300	0.0037	0.0134	0.0097	2.6300	0.0132	0.0120	-0.0012	-0.0900
C8	8	0.0012	0.0013	0.0001	0.0900	0.0127	0.0254	0.0127	1.0100	0.0166	0.0013	-0.0153	-0.9200
C9	7	0.0011	0.0036	0.0025	2.2300	0.0111	0.0277	0.0166	1.5000	0.0097	0.0036	-0.0062	-0.6300
C10	5	0.0148	0.0048	-0.0100	-0.6700	0.0085	0.0278	0.0193	2.2700	0.0086	0.0048	-0.0038	-0.4400
AVG		0.0030	0.0004	0.0007	0.0004	108%	0.00623	0.0227	0.01643	264%	0.0068	0.0007	-0.0060

4.4 Efficiency Analysis

The efficiency analysis presented in Table 4.4 evaluates the performance of three applications—Farah App, Dar Al Consul (DAC), and EasyApp—across two sessions, focusing on both time-based efficiency and overall relative efficiency. These metrics provide insights into each application's ability to facilitate quick and efficient task completion over repeated use.

Farah App: For Farah App, time-based efficiency improved significantly from Session 1 (0.4675) to Session 2 (0.5561), with an increase of 0.0886. This positive change suggests that users became more adept at completing tasks quickly with repeated interaction, indicating a high degree of learnability and efficiency gain over time. However, the overall relative efficiency for Farah App decreased from 0.0021 in Session 1 to a very low 0.00003 in Session 2, with a substantial drop of

-0.3190. This discrepancy suggests that while users became faster, the overall efficiency may have been impacted by factors such as increased task complexity or additional steps that affected the relative efficiency metric

Table 4.6 Efficiency analysis of Farah, DAC and EasyApp applications using the automated framework

APP	Time Based Efficiency			Overall relative Efficiency		
	Session 1	**Session 2**	**Difference**	**Session 1**	**Session 2**	**Difference**
Farah App	0.4675	0.5561	0.0886	0.0021	0.00003	-0.3190
Dar AL Consul App	0.0444	0.0414	-0.0030	0.0007	0.0011	0.0003
EasyAPP	0.0357	0.0318	-0.0039	0.0006	0.0005	-0.0001
Average	0.1825	0.2098	0.0272	0.1077	0.0014	-0.1062

Dar Al Consul (DAC): DAC shows a slight decline in time-based efficiency, dropping from 0.0444 in Session 1 to 0.0414 in Session 2, with a difference of -0.0030. This decrease suggests that users did not improve in terms of task completion speed, possibly due to the application's interface complexity or challenging task flow that did not support efficiency gains over time. However, the overall relative efficiency for DAC showed a minor improvement, increasing from 0.0007 in Session 1 to 0.0011 in Session 2, with a difference of 0.0003. This minor improvement in relative efficiency could indicate that users were able to perform certain tasks with more precision, even if the overall speed did not increase.

EasyApp: EasyApp exhibited a minor decrease in time-based efficiency, going from 0.0357 in Session 1 to 0.0318 in Session 2, with a difference of -0.0039. This reduction in efficiency indicates that users either encountered challenges that hindered quick task completion or required more steps to complete tasks in Session 2. Similarly, overall relative efficiency slightly declined from 0.0006 to 0.0005, with a difference of -0.0001. This suggests that EasyApp's efficiency was relatively stable but did not improve over time, potentially highlighting areas where the interface or task structure could be optimized for better user efficiency.

On average, there was a modest improvement in time-based efficiency across all applications, from 0.1825 in Session 1 to 0.2098 in Session 2, with an overall increase of 0.0272. This suggests that users generally became slightly faster at task completion across sessions. However, the average overall relative efficiency decreased significantly, from 0.1077 to 0.0014, with a notable drop of -0.1062. This decline points to a potential trade-off where increased speed may not have correlated with improved accuracy or optimal task flow, particularly for Farah App, which saw the largest drop in relative efficiency. Furthermore, while time-based efficiency improvements indicate that users became faster with repeated use, the mixed results in overall relative efficiency suggest that these gains may not fully translate to improved overall task effectiveness. Farah App showed the highest improvement in time-based efficiency but a substantial decline in relative efficiency, suggesting areas where task flow could be streamlined. DAC and EasyApp maintained relatively stable efficiencies but showed little overall improvement, indicating a need for interface or user experience enhancements to support more efficient task performance over time. This can be enhanced by clearly defining tasks and organizing them into a sequential flow. Additionally, improving the user interface by incorporating intuitive gestures, visual indicators, and instructional pointers can help guide users more effectively and streamline task completion.

4.4 Memorability Analysis

The memorability analysis in Table 4.5 evaluates three applications—Farah App, Dar Al Consul (DAC), and EasyApp—using several key metrics: Task Retention Score (TRS), Time to Relearn (TTR), Error Rate (ER), and an overall memorability score (M). This analysis helps assess how well users retained knowledge and proficiency over time for each application, with higher scores indicating better memorability.

Table 4.7 Memorability analysis of Farah, DAC and EasyApp applications using the automated framework

User	Farah App				Dar AlConsul App				EasyAPP			
	(TRS)	(TTR)	ER	M1	(TRS)	(TTR)	ER	M2	(TRS)	(TTR)	ER	M3
A1	1.00	3.01	0.00	0.3322	1.00	0.53	0.00	1.8985	1.00	0.68	0.00	1.4608
A2	0.67	3.12	0.00	0.2138	1.00	0.54	0.00	1.8574	1.00	0.43	0.00	2.3295
A3	0.67	1.96	0.33	0.2905	1.00	0.54	0.00	1.8429	1.00	0.51	0.00	1.9640
A4	1.00	3.32	0.33	0.2737	1.00	0.64	0.00	1.5613	1.00	0.51	0.00	1.9508
A5	0.67	2.74	0.33	0.2172	1.00	0.61	0.00	1.6468	1.00	2.45	0.00	0.4089
A6	0.67	3.44	0.00	0.1939	1.00	1.06	0.00	0.9454	1.00	1.02	0.00	0.9781
T1	1.00	6.46	0.00	0.1547	0.67	0.49	0.33	0.8130	1.00	0.49	0.00	2.0462
T2	0.67	4.00	0.00	0.1666	0.67	0.34	0.33	0.9992	1.00	0.75	0.00	1.3249
T3	1.00	4.00	0.00	0.2497	1.00	0.43	0.00	2.3178	1.00	0.65	0.00	1.5300
C1	1.00	4.35	0.00	0.2298	1.00	0.57	0.00	1.7683	1.00	0.57	0.00	1.7540
C2	0.67	2.99	0.00	0.2233	1.00	0.77	0.00	1.2954	1.00	0.58	0.00	1.7356
C3	1.00	4.17	0.00	0.2398	1.00	0.76	0.00	1.3101	1.00	0.53	0.00	1.8973
C4	1.00	5.88	0.00	0.1700	1.00	0.59	0.00	1.6818	1.00	0.55	0.00	1.8220
C5	1.00	3.34	0.00	0.2992	1.00	0.80	0.00	1.2473	1.00	0.60	0.00	1.6736
C6	1.00	4.42	0.00	0.2262	1.00	1.28	0.00	0.7835	1.00	1.12	0.00	0.8933
C7	0.67	0.88	0.33	0.5499	1.00	1.02	0.00	0.9830	1.00	1.39	0.00	0.7215
C8	0.33	2.46	0.33	0.1194	1.00	0.62	0.00	1.6213	0.33	1.39	0.67	0.1618
C9	0.33	0.73	0.00	0.4535	1.00	0.46	0.00	2.1900	0.67	2.08	0.33	0.2760
C10	1.00	2.66	0.66	0.3016	1.00	0.36	0.00	2.7778	0.67	1.54	0.33	0.3564

Farah App: Users displayed variable memorability for Farah App. Task Retention Scores (TRS) were high for most users, with many achieving a perfect score of 1.00, indicating that users could retain task-related knowledge effectively. However, Time to Relearn (TTR) values varied significantly, ranging from around 0.73 to as high as 6.46. Higher TTR values, such as 6.46 for T1, indicate that certain users took longer to regain proficiency, suggesting some areas where the app's interface could be optimized to improve learnability. Error Rates (ER) for Farah App were generally low, with most users achieving zero errors, which positively impacted the overall memorability score (M1). Memorability scores (M1) ranged from 0.1194 for user C8 to 0.5499 for C7, indicating moderate overall memorability.

Dar Al Consul (DAC): The DAC app demonstrated high memorability across most metrics. Most users achieved a TRS of 1.00, indicating strong task retention. Time to Relearn (TTR) values were relatively low for DAC, with most users averaging under 1.0, indicating that users could quickly reacquaint themselves with tasks. Error Rates (ER) remained at zero for nearly all users, contributing to consistently high memorability scores (M2). The memorability scores for DAC ranged from 0.7835 for C6 to 2.7778 for C10, suggesting that users generally found DAC easy to remember and quickly regained proficiency. This indicates that the DAC app offers a highly user-friendly interface that supports efficient task recall and interaction.

EasyApp: EasyApp also exhibited high memorability, with most users scoring 1.00 in TRS, suggesting good retention of task steps and knowledge. However, TTR scores varied more widely than for DAC, with some users, such as A5 and C7, showing higher TTR values, indicating that it took them longer to reattain their initial proficiency level. Error Rates (ER) were minimal, although user C8 exhibited a higher error rate of 0.67, which slightly impacted memorability. Memorability scores (M3) for EasyApp ranged from 0.1618 for C8 to 2.3295 for A2, showing that while memorability was high overall, there were isolated instances where task recall required additional time or incurred more errors.

Across all three applications, memorability was generally high, as evidenced by high TRS scores and low error rates for most users. Farah App displayed the most variation in TTR, suggesting that while task retention was strong, some users took longer to return to their previous proficiency levels. DAC consistently showed high memorability scores across all users, with low TTR and ER, indicating an efficient and intuitive interface that aids user recall and proficiency. EasyApp showed high memorability overall, although certain users exhibited longer relearning times and isolated errors, suggesting that minor adjustments could improve its user retention.
Moreover, the DAC app demonstrated the highest overall memorability, followed closely by EasyApp, while Farah App showed moderate memorability with some opportunities for improvement in user proficiency retention. This analysis underscores the importance of a user-friendly interface and intuitive task flow in promoting high memorability across different user demographics and applications.

4.5 Quantitative findings

System Usability Scale (SUS) was administered to all participants at the end of their sessions. The SUS analysis in Table 4.8 revealed that Farah App had the highest average score (0.7505) and the most consistent usability ratings, indicating a stable and positive user experience. DAC App scored slightly lower (0.7085) but exhibited a wide range of ratings (0.37 to 0.97), suggesting polarized user experiences with some users finding it very usable while others struggled. EasyApp had the lowest average score (0.6825) and the highest variability, reflecting significant usability challenges for some users despite achieving the highest individual score (0.98). These results highlight Farah App's strength in usability, while DAC App and EasyApp require targeted improvements to enhance consistency and overall user satisfaction.

Table 4.8: SUS scores for tested applications.

SUS score (0-1)	A1	A2	A3	A4	A5	A6	T1	T2	T3	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Farah App	0.84	0.93	0.92	0.69	0.81	0.76	0.94	0.82	0.75	0.83	0.90	0.75	0.69	0.91	0.82	0.94	0.76	0.60	0.73
DAC App	0.96	0.93	0.86	0.78	0.83	0.63	0.73	0.37	0.47	0.82	0.69	0.79	0.46	0.76	0.63	0.57	0.92	0.97	0.65
EasyApp	0.71	0.82	0.43	0.68	0.87	0.95	0.92	0.60	0.51	0.98	0.48	0.58	0.98	0.70	0.72	0.66	0.67	0.93	0.92

The QUIS questionnaire in Table 4.9 revealed the usability evaluation of three AR applications—Farah App, Dar Al Consul (DAC), and Easy AR. Farah App achieved the highest overall satisfaction score (4.65/5), with users highlighting its engaging tasks, intuitive learning experience, and visually appealing interface. DAC followed with an average score of 4.45/5, excelling in visual quality and learning, though it faced minor challenges with task complexity, such as "Rotating Marbles." Easy AR scored the lowest (4.32/5), primarily due to a less intuitive GUI design and occasional issues with tasks like "Resizing Object."

Table 4.9: QUIS Scores for the Tested Applications.

App / Section	Overall Satisfaction	Visual Quality	GUI	Terminology & System Info	Learning	System Capabilities	Average Satisfaction Score
Farah App	4.5	4.7	4.2	4.8	4.9	4.8	4.65
DAC	4.2	4.5	4	4.6	4.7	4.7	4.45
Easy AR	4	4.4	3.8	4.5	4.6	4.6	4.32

Despite strong system capabilities and learning experiences across all three applications, Farah App stood out as the most user-friendly and engaging, while Easy AR requires improvements in interface design and task responsiveness.

Chapter 5:

Discussion and Conclusion

5.1. Discussion

This study applied the automated measure framework to evaluate three AR applications—Farah App, Dar Al Consul (DAC), and EasyApp—focusing on learnability, effectiveness, efficiency, and memorability to understand how each app performs in various contexts, including educational, tourism, and architectural settings. The findings contribute to the broader discourse on AR usability and emphasize the need for age- and task-specific design adaptations.

The learnability analysis demonstrated considerable variation across applications and age groups. Farah App, for instance, showed a high learnability rate among younger users, with participants aged 8 and 6 displaying ease in adapting to the tasks and improving their performance over time. This is consistent with studies indicating that younger users tend to be more adaptable to digital environments, especially when the interfaces incorporate elements of gamification or experiential learning (Milgram et al., 2011). Farah App's interface seemed to align well with the cognitive and motor capabilities of younger users, facilitating a smoother learning curve. However, older users encountered moderate learnability scores, indicating that age significantly impacts ease of adaptation in AR environments, supporting findings by Rasheed et al. (2021), who highlighted the role of age-related cognitive and motor differences in digital adaptability.

For DAC and EasyApp, the learnability scores were relatively uniform across age groups, suggesting these applications offer a more balanced user experience that supports diverse age ranges. This aligns with Azuma's (2016) observation that standardized, intuitive navigation paths can mitigate age-related usability differences, allowing users of varying cognitive capabilities to adapt quickly. However, some users, like T1 and A6, exhibited negative learnability scores in DAC, which could be attributed to the app's interface complexity or task structure, which demanded more initial cognitive load. Bowman et al. (2012) have similarly noted that when AR interfaces lack clear task cues, users may experience higher cognitive strain, particularly in early stages, underscoring the need for interface simplification to support rapid adaptation.

In terms of effectiveness, measured through task completion rates, error rates, and accuracy, the results varied significantly across applications. Farah App displayed an increase in task completion rates and a reduction in error rates over time, suggesting that users became more proficient with continued exposure. This trend is consistent with the theory of repeated exposure leading to greater proficiency in digital interactions, as shown by Dünser et al. (2008), where repeated engagement with AR interfaces improved performance metrics and reduced user errors. Conversely, DAC showed a slight decrease in task completion rates coupled with a substantial increase in error rates from 0.016 to 0.333, which may point to underlying challenges in the application's interface complexity. Azuma (2017) highlights that AR applications with higher cognitive loads tend to elevate error rates, indicating that DAC may require modifications to simplify navigation or task structures. EasyApp maintained high accuracy yet showed a slight decline in task completion, which could suggest minor usability issues that could potentially be addressed through optimized task flows or enhanced instruction clarity.

The Overall Task Success Rate (OTSR) analysis provided further insight into the role of age in task success across applications. Farah App demonstrated significant improvements in OTSR among younger users, with increases of 2345% and 837% for users aged 8 and 6, respectively. This marked improvement indicates that the application's design is well-suited to the cognitive and interactive preferences of younger users, corroborating findings by Rasheed et al. (2021), which show that gamified AR applications in educational contexts improve engagement and task success for children. In contrast, DAC presented mixed results, with some older users demonstrating task improvements while younger users exhibited declines, highlighting usability challenges, especially for children. Gabbard and Swan (2008) identified that younger users often struggle with interfaces that require fine motor control or complex navigation, and this could explain the decrease in OTSR for DAC among younger participants. EasyApp generally showed OTSR improvements, although declines for younger participants suggest that even in relatively well-designed interfaces, specific usability enhancements may be required to accommodate very young users.

Efficiency analysis, focusing on time-based efficiency and overall relative efficiency, revealed both gains and trade-offs. Farah App displayed a notable increase in time-based efficiency, suggesting that users became faster at task completion with repeated interactions. However, the app's overall relative efficiency decreased sharply, suggesting that faster task completion was not necessarily accompanied by smoother task flow. This phenomenon of increased speed but diminished task flow efficiency has been observed by Bowman and McMahan (2007), who found that simplified task flows can enhance speed but may inadvertently add cognitive load, especially when tasks require additional steps. For DAC and EasyApp, time-based efficiency remained consistent but relatively low, indicating stable but unspectacular task completion speeds. This aligns with prior findings that suggest moderate task complexity can support consistent user interaction but may not drive increased efficiency without further interface adjustments (Schuemie et al., 2001).

The memorability analysis, which assessed Task Retention Scores (TRS), Time to Relearn (TTR), and Error Rates (ER), showed that users generally retained task knowledge across sessions. Farah App displayed high TRS scores but significant variability in TTR, indicating that while task retention was strong, some users required additional time to regain proficiency. This could point to a need for more intuitive task reintroduction, a factor emphasized by Hassenzahl and Tractinsky

(2006), who found that clear task cues in AR interfaces improve user retention. DAC users demonstrated high memorability, characterized by low TTR values and consistently high TRS scores, reflecting a user-friendly interface that facilitates task recall. Garrett (2010) supports this observation, noting that well-streamlined interfaces with intuitive navigation bolster user memory and retention. EasyApp similarly displayed high memorability scores, although certain users exhibited prolonged relearning times, likely due to specific tasks' complexity.

Comparing these findings with related studies in AR usability underscores the importance of age-tailored adjustments, as the variance in learnability, efficiency, and memorability scores across age groups and applications highlights the need for user-centered design that accommodates cognitive and motor skills unique to each demographic. Norman (2013) emphasizes that simplicity and user-centered design are critical in AR systems, as they support intuitive interaction and reduce cognitive strain, particularly for less experienced users. The results also affirm the importance of low-complexity navigation for improving efficiency and accuracy over time.

In conclusion, the study's findings, corroborated by existing research, indicate that while AR applications can achieve high learnability and memorability, developers must balance task complexity with speed and accuracy to ensure overall efficiency. The analysis suggests that applications like DAC, which prioritize streamlined navigation, demonstrate the highest memorability scores, while applications catering to younger users, such as Farah App, perform exceptionally well in learnability and task success. These results form a foundation for future enhancements, suggesting that developers should prioritize age-appropriate interactions, simplify task flows, and incorporate clear, intuitive feedback mechanisms to maximize user experience and retention in AR environments.

The study's findings emphasize the need for a standardized AR user experience framework that evaluates both learnability and cognition in AR products, considering factors such as age, task complexity, and feedback clarity. By employing the VARTOOL framework, this study contributes to a growing body of knowledge on automated AR usability testing and offers a versatile tool for developers. VARTOOL, validated across three application types—education, tourism, and architecture—proved effective in measuring learnability metrics and can serve as a basis for further research and development in AR usability, ultimately enhancing user engagement, reducing error rates, and improving task efficiency across diverse AR applications.

5.2. Conclusion

The study sought to evaluate the learnability, effectiveness, efficiency, and memorability of three augmented reality (AR) applications—Farah App, Dar Al Consul (DAC), and EasyApp—across educational, tourism, and architectural contexts using the automated framework. Through a comprehensive analysis of user interactions across diverse age groups, the study provides critical insights into how AR applications perform across these dimensions, offering actionable data for developers aiming to enhance user experience in AR environments.

The findings demonstrate that each application's performance varied depending on user demographics and application purpose, with Farah App showing high learnability and task adaptability among younger users, aligning well with educational contexts where experiential learning is essential. This supports the theory that age-specific design adjustments can significantly improve user experience in AR applications, particularly for children and non-technical users. Conversely, DAC and EasyApp displayed consistent usability across all age groups, indicating their potential to serve broader audiences but also highlighting the need for interface simplifications to further improve task efficiency and reduce error rates.

Efficiency and memorability analyses provided additional insights, revealing a trade-off between speed and task complexity across applications. While Farah App showed improvement in time-based efficiency, its overall relative efficiency decreased, underscoring that faster task completion does not necessarily correlate with optimal navigation and usability. DAC and EasyApp maintained stable, though moderate, efficiencies, with DAC displaying the highest memorability scores due to a streamlined interface and intuitive navigation. The memorability findings confirm that clear task cues and simplified workflows are crucial in supporting task retention and quick reacquaintance in AR interfaces.

The study's findings are consistent with existing literature emphasizing the importance of age-appropriate interfaces, simplicity in navigation, and intuitive feedback mechanisms for improving user retention and engagement in AR systems. The framework proved effective in capturing these insights, validating its potential as a reliable automated tool for evaluating and enhancing AR usability metrics. Furthermore, the findings demonstrate that the proposed automated framework significantly enhances the efficiency and effectiveness of user experience testing in AR applications by reducing the time, effort, and resources required. The integration of Unity and Firebase enabled real-time data collection and automated analysis of key metrics such as task completion rate, error rate, and user satisfaction, eliminating the need for manual observation and minimizing human bias. The framework also streamlined testing processes, with VARTOOL dynamically displaying results on dashboards, allowing developers to identify issues and implement improvements more quickly. Additionally, by automating workflows and reducing reliance on manual surveys and observations, the framework optimized resources and lowered costs associated with traditional usability testing methods. These advancements confirm the framework's ability to accelerate and simplify UX testing in AR applications.

The framework's accuracy in assessing UX in AR applications, particularly task completion, was validated through comprehensive task completion metrics, cross-application comparisons, and user feedback. It measured task completion rates, error rates, and task accuracy, using the Overall Task Success Rate (OTSR) as a composite metric that combined task completion rate, average completion time, and error rate for a holistic evaluation. Consistent results across different AR applications, such as higher learnability and task success for children using the Farah App, demonstrated the framework's ability to assess UX accurately for specific demographics. Additionally, user satisfaction data from SUS questionnaires confirmed that the framework's automated assessments aligned closely with subjective user experiences. These findings highlight the framework's reliability in capturing both objective and subjective aspects of UX in AR applications.

The study highlights the framework's significant contributions to enhancing UX quality in AR applications through actionable, data-driven insights. It improved user experience by tracking metrics such as error reduction over time and time to task proficiency, enabling the identification of areas for improvement in UI design and task flows. The framework also provided personalized insights into user interactions, helping developers tailor applications to specific user groups, such as children and adults, to better meet their needs. By incorporating a multidimensional assessment of UX—spanning learnability, memorability, effectiveness, efficiency, and satisfaction—the framework allowed developers to enhance both functional usability and emotional engagement. These contributions underscore the framework's impact in driving iterative design improvements and elevating overall UX quality.

In conclusion, this study highlights that while AR applications offer substantial potential for immersive and engaging user experiences, there is a critical need for targeted UX optimizations tailored to the unique needs of diverse user groups and application contexts. Future development in AR should prioritize user-centered design principles, focusing on task flow simplicity, age-specific adaptability, and improved efficiency without compromising memorability. These considerations will be essential in leveraging AR's full potential across educational, tourism, architectural, and other application domains, contributing to a more inclusive and impactful digital experience. The framework offers a strong foundation for ongoing research, providing developers with a robust tool to assess and optimize learnability and associated usability factors in AR applications.

5.3. Future Work

Refinement of Usability Metrics: Future work should focus on refining the usability metrics within the framework to provide even more granular insights into user interactions. This could include developing new metrics that capture the nuances of user experience in AR environments, such as cognitive load or emotional response.

Extended Testing Across Diverse User Groups: To further validate the tool, future studies should involve a broader demographic, including users with varying levels of AR experience, different age groups, and those with cognitive or physical impairments. This would help in identifying specific challenges faced by different user segments and inform more inclusive design practices.

Longitudinal Studies: Conducting longitudinal studies that track user performance over an extended period could provide deeper insights into how users adapt to AR environments over time. This would also help in assessing the long-term learnability and retention of tasks within AR applications.

Integration with Other Usability Tools: Integrating framework with other usability testing tools or platforms could enhance its capabilities, allowing for a more comprehensive evaluation of AR applications. This could include combining framework with eye-tracking or biometric analysis tools to assess user engagement and cognitive load.

Exploration of New Application Domains: Expanding the scope of framework to test AR applications in new domains, such as healthcare, retail, or entertainment, could provide insights into the specific usability challenges and requirements of these sectors. This would contribute to the development of best practices for AR usability across different industries.

Adaptive Learning and Personalization: Future enhancements to the tested AR applications could include the development of adaptive learning features that personalize the user experience based on individual performance. This could help users who struggle with certain tasks by providing tailored guidance or training, ultimately improving overall usability.

These future directions will not only strengthen the findings of this study but also contribute to the broader field of AR usability, ensuring that AR applications are accessible, effective, and user-friendly for a wide range of users.

References

1. Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments/MIT Press*.
2. Rasheed, G., Khan, M., Malik, N., & Akhunzada, A. (2021). Measuring learnability through virtual reality laboratory application: A user study. *Sustainability (Switzerland)*, *13*(19). https://doi.org/10.3390/su131910812
3. Thompson, S. K., & Patel, R. M. (2017). Tracking and performance in augmented reality systems: A survey. *ACM Transactions on Graphics*, *36*(6), 1–20.
4. Johnson, C. P., & Lee, M. R. (2021). Real-time performance analysis in augmented reality applications. *International Journal of Computer Vision*, *129*(4), 789–804.
5. Chen, P. , L. X. , C. W. , & H. R. (2020). A review of augmented reality in education: Comparisons of the literature from 2011 to 2019. *Computers & Education*, 103–707.
6. Hughes, C. , et al. (2017). Augmented reality in learning and teaching. In Handbook of Augmented Reality. *Springer, New York.*, 667–685.
7. Anderson, M. J., & Ramirez, P. C. (2018). Performance evaluation of augmented reality systems: An empirical approach. *IEEE Transactions on Visualization and Computer Graphics*.
8. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34–47.
9. Billinghurst, M., Clark, A., & Lee, G. (2014). A survey of augmented reality. In *Foundations and Trends in Human-Computer Interaction* (Vol. 8, Issues 2–3, pp. 73–272). Now Publishers Inc. https://doi.org/10.1561/1100000049
10. ISO 9241-210. (2019). *Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems*. International Organization for Standardization.
11. Smith, A. L., & Thomas, C. J. (2020). The impact of user experience on augmented reality performance: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, *26*(7), 1–12.
12. Schneider, M. B., & Liu, E. F. (2018). Evaluating user experience in augmented reality: A framework for analysis. *ACM Transactions on Computer-Human Interaction*, *25*(4), 1–26.
13. Reinders, J. S., & Williams, L. K. (2021). Optimizing AR performance: Frame rate and latency considerations. *International Journal of Computer Vision*, *129*(5), 1574–1592.
14. Dix, A. , F. J. , A. G. D. , & B. R. (2004). Human-computer interaction. *Pearson Education*.
15. B. Shneiderman and C. Plaisant. (2010). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. http://www.aw-bc.com/dtui.c1icking
16. Hassenzahl, M., & Tractinsky, N. (2006). User experience - A research agenda. *Behaviour and Information Technology*, *25*(2), 91–97. https://doi.org/10.1080/01449290500330331
17. Norman Donald, A. (2013). *The design of everyday things*. MIT Press.
18. Nielsen, Jakob., & Budiu, Raluca. (2013). *Mobile usability*. Pearson Education.
19. Preece et al. (2015). *Interaction Design: Beyond Human-Computer Interaction*.

20. Albert, B., & Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes.
21. Kim, S. , et al. (2018). User experience evaluation of AR and VR headsets. *Journal of Virtual Reality and Broadcasting*, 1–9.
22. Craig, A. B. (2013). *Understanding augmented reality: Concepts and applications*.
23. Hartson, R., & Pyla, P. S. (2012). *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.
24. Liu, D. , et al. (2020). How augmented reality supports students' learning in STEM: A meta-analysis. *Educational Research Review*, 100–101.
25. Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in Psychology*, *9*, 2086.
26. Chen, P. , et al. (2017). The application of augmented reality for the training of maintenance operators in industrial environments. *Procedia Manufacturing*, 1255–1262.
27. Rauschnabel, P. A., Rossmann, A., & tom Dieck, M. C. (2017). An adoption framework for mobile augmented reality games: The case of Pokémon Go. *Computers in Human Behavior*, *76*, 276–286.
28. Wojciechowski, R., & Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers & Education*, *68*, 570–585.
29. Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. *Handbook of Research on Educational Communications and Technology*, 735–745.
30. Tussyadiah, I. P., Wang, D., Jung, T. H., & Tom Dieck, M. C. (2018). Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management*, *66*, 140–154.
31. Jordan, P. W. (2000). *Designing pleasurable products: An introduction to the new human factors*. CRC press.
32. Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann. https://books.google.ps/books?id=95As2OF67f0C&lpg=PR9&ots=3dAECoaoXx&dq=usability%20engineering%20nielsen&lr&pg=PR11#v=onepage&q=usability%20engineering%20nielsen&f=false
33. Johnson, K. , et al. (2021). Evaluating learnability in augmented reality applications. In *Computers & Education*.
34. Theng, Y. L., Nasir, N. M., Thimbleby, H., Buchanan, G., Jones, M., Bainbridge, D., & Cassidy, N. (2000). Children as design partners and testers for a children's digital library. *Research and Advanced Technology for Digital Libraries: 4th European Conference, ECDL 2000 Lisbon, Portugal, September 18–20, 2000 Proceedings 4*, 249–258.
35. Hartson, R., & Pyla, P. S. (2012). *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.
36. Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. *Handbook of Research on Educational Communications and Technology*, 735–745.
37. Panagiotis, D. (2011). Evaluating UX in AR and VR: A Framework for Enhanced Interactivity. *Presence: Teleoperators & Virtual Environments,* , 290–305.

38. Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.
39. International Organization for Standardization (ISO). (2011). *ISO 25010 2011*.
40. Steed, A., & Oliveira, M. F. (2009). *Networked graphics: building networked games and virtual environments*. Elsevier. https://books.google.ps/books?id=76C_quJqVXcC&lpg=PP1&ots=RZ3xulSfFQ&dq=Networked%20graphics%3A%20building%20networked%20games%20and%20virtual%20environments&lr&pg=PR4#v=onepage&q=Networked%20graphics:%20building%20networked%20games%20and%20virtual%20environments&f=false
41. Gabbard, J. L., & Swan II, J. E. (2008). Usability engineering for augmented reality: Employing user-based studies to inform design. *IEEE Transactions on Visualization and Computer Graphics*, *14*(3), 513–525.
42. Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer*, *40*(7), 36–43.
43. Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. P. G. (2001). Research on presence in virtual reality: A survey. *Cyberpsychology & Behavior*, *4*(2), 183–201.
44. Garrett, J. J. (2010). *The elements of user experience: User-centered design for the web and beyond*.
45. Schumann, M., Fuchs, C., Kollatsch, C., & Klimant, P. (2020). Evaluation of augmented reality supported approaches for product design and production processes. *Procedia CIRP*, *97*, 160–165. https://doi.org/10.1016/j.procir.2020.05.219
46. Gao, Y. , Z. T. , & L. Y. (2018). Augmented Reality for Enhancing Product Design and Prototyping. *International Journal of Product Development*.
47. Pérez, S. , & M. M. (2021). Enhancing Prototyping Accuracy and Efficiency with Augmented Reality in Product Design. *Journal of Product Innovation Management*, 112–130.
48. Kim, H. , & L. J. (2017). User-Centered Design Strategies in AR for Product Development. *Design Journal*, 152–165.
49. Grasset, R., Duenser, A., Seichter, H., & Billinghurst, M. (2007). The mixed reality book: a new multimedia reading experience. *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, 1953–1958.
50. Henderson, S. , & F. S. (2011). Exploring the Benefits of AR for Supporting User-Centered Design and Prototyping. In S. , & F. S. Henderson (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 465–472).
51. Guerra-Filho, G. , & G. R. (2020). Adapting User-Centered Design Principles for Augmented Reality Applications. *Journal of Human-Computer Studies*, 65–80.
52. Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web: methods and strategies of web personalization* (pp. 3–53). Springer.
53. Baird, D. , & M. J. (2018). (n.d.). Applying User-Centered Design in Augmented Reality Development: A Case Study. *International Journal of Human-Computer Interaction*, 254–270.
54. Vargas, Y. A. M., Rojas, E. E. M., & Castillo, V. S. (2019). Application of augmented reality as a means of interdisciplinary learning. *Educational Research and Reviews*, *24*(3), 375–386.

55. Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, *123*, 109–123.
56. Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, *18*, 7–22.
57. Huang, R. , et al. (2021). Augmented Reality in Corporate Training: Effects on Learning Speed and Retention. *International Journal of Educational Technology in Higher Education*, 18–39.
58. Grossman, T. (2009). A Study of Learnability: Classification and Improvement Techniques in Software Interfaces. *ACM Transactions on Computer-Human Interaction*.
59. Blandford, A. , G. T. (2008). From Work to Study: How Work-Focused Learning Can Inspire Understanding in Software Usability and Learnability. *International Journal of Human-Computer Studies*, *66*(4), 252–271.
60. Hafsa, S. (2021). Developing a Learnability Enhancement Model for AR Interfaces. *Journal of Human-Computer Interaction*, *37*(6), 453–478.
61. Kron, F. W. , L. C. (2016). Assessing Learnability in Augmented Reality Applications: Challenges and Metrics. *Computers Education*, *95*, 110–120.
62. Cheng, K. H. , & T. C. C. (2013). Affordances of Augmented Reality in Science Learning: Suggestions for Future Research. *Journal of Science Education and Technology*, *22*(4), 449–462.

إطار عمل لتجربة مستخدم الواقع المعزز: قياسات مؤتمتة لتجربة المستخدم والعوامل المرتبطة بها

اعداد : بلقيس احمد خليل عواوده

اشراف : د. رضوان القصراوي

الملخص:

الخلفية: تتطور تكنولوجيا الواقع المعزز (AR) بسرعة، مما يفتح فرصًا جديدة لتجارب المستخدم (UX)عبر مجالات مثل التعليم والسياحة والهندسة المعمارية ومع ذلك، فإن تطوير أطر عمل آلية تقوم بتقييم تجربة المستخدم بشكل شامل أمر بالغ الأهمية، خاصة من حيث قابلية التعلم، والكفاءة، والفعالية، وقابلية التذكر، استنادًا إلى معايير التفاعل الموجهة نحو الإنسان المعتمدة.
الهدف: تهدف هذه الدراسة إلى تطوير إطار عمل لاختبار تجربة المستخدم في تطبيقات الواقع المعزز، مع التركيز الأساسي على تقييم مقاييس قابلية الاستخدام والمشاركة من قبل المستخدمين. يعتمد الإطار على مبادئ التفاعل بين الإنسان والحاسوب المعتمدة كأساس للإجراءات الآلية، مما يوفر أداة تقييم شاملة وموضوعية لدعم تطوير واجهات واقع معزز بديهية وسهلة الاستخدام تعزز من تفاعل المستخدم.
الطريقة: تم استخدام الإطار التلقائي لتقييم سهولة التعلم والتفاعل والكفاءة عبر ثلاث تطبيقات للواقع المعزز – تطبيق فرح (التعليمي)، دار القنصل (السياحي)، و EasyApp (الهندسي) – مع مجموعة مكونة من 20 مستخدمًا يمثلون فئات عمرية متنوعة. تم جمع وتحليل بيانات التفاعلات مع المستخدمين، كما تم جمع تعليقات المستخدمين عبر استبيانات لتكملة التحليل التلقائي.
النتائج :أظهرت النتائج اختلافات ملحوظة في سهولة الاستخدام والفعالية بين التطبيقات والمستخدمين. أظهر تطبيق فرح، على وجه الخصوص، مستويات عالية من التعلم السهل، خاصة بين المستخدمين الأصغر سنًا، مما يعكس تميز التطبيق كأداة تعليمية متناسبة مع الفئة العمرية. أما تطبيق دار القنصل وEasyApp، فقد أظهرا استقرارًا في سهولة الاستخدام عبر الفئات العمرية، مع بعض النقاط التي يمكن تحسينها في الكفاءة وتقليل الأخطاء.
الخاتمة :يؤكد هذا البحث على أهمية تطوير أطر عمل تعتمد على معايير تفاعل الإنسان مع الحاسوب لتقييم تجربة المستخدم في تطبيقات الواقع المعزز. وتشير النتائج إلى أن إطار العمل المطور يوفر قياسات شاملة وموضوعية لتعزيز تجربة المستخدم، مع توصيات لتصميم واجهات ملائمة لجميع الأعمار وتبسيط التدفق التفاعلي لضمان تجربة استخدام فعّالة ومستدامة عبر مختلف تطبيقات الواقع المعزز.