ISSN 2723-6471

Mixed Method Usability Testing for User Experience and User Interface of AI-Based Supermarket Applications

Esti Dwi Rinawiyanti[1,*], Rosita Meitha Surjani[2], Markus Hartono[3],
Eunike Putri Permatasari[4], Angela Carla Vania Chrisma[5]

[1,2,3,4,5]*Study Program of Industrial Engineering, Faculty of Engineering, University of Surabaya, Surabaya 60239, Indonesia*

(Received: June 30, 2024; Revised: September 13, 2024; Accepted: October 17, 2024; Available online: December 30, 2024)

Abstract

Designing digital apps for businesses, including retail companies, is crucial in the digitalization era. To enhance the customers' shopping experience, an AI-based digital app for supermarkets has been designed to facilitate customers' searching for and purchasing goods. AI-based digital apps with a good user interface and user experience design will be very effective because they can enhance the app's visual appeal. After developing the digital app based on the user's needs and preferences, usability testing was conducted to assess whether it could overcome the problem and enhance the shopping experience. This study aims to determine the usability of digital apps using mixed methods through qualitative and quantitative methods. The usability testing was conducted quantitatively through Maze software (for seven tasks) and a questionnaire using Nielsen's aspects, including learnability, efficiency, memorability, accuracy, and satisfaction. The results of the Maze score indicated an 80% success rate and a usability score of 59%. This result refers to a medium usability score, indicating the need for future improvements. The results from the usability testing using the questionnaire showed high scores on learnability (85%), efficiency (80%), and memorability (83%), as well as medium scores on accuracy (60%) and satisfaction (50%). The average score from questionnaires is 72%. Both results implied medium usability scores. After quantitative testing, qualitative data regarding users' experiences, problems, and expectations using the digital app were collected through interviews and observations. The results of the quantitative and qualitative analysis show that the digital shopping app should improve its layout, ease of use, features like a favorites section, and product descriptions. The findings of this study offer fresh insight into integrating quantitative and qualitative methods. In contrast, quantitative and qualitative testing results can be combined to provide a comprehensive analysis that will help the apps improve.

Keywords: UI/UX, Digital App, Usability Testing, Mixed Method

1. Introduction

In 2022, the retail sales value of supermarkets in Indonesia totaled approximately 4.93 billion USD, an increase of about 0.6 billion USD compared to the previous year [1]. Supermarkets offer various products, and each category requires a distinct arrangement. According to observations and interviews, consumers frequently need help finding the products they wish to purchase because there are no instructions outlining the locations of the goods. This difficulty will certainly affect consumers' comfort in shopping.

Nonetheless, this issue generates an opportunity to create digital apps that make it easier for consumers to shop offline. The apps make the shopping process more effective and time-efficient, remind consumers of special offers, provide recommendations based on shopping preferences, and display frequently purchased goods. Thus, the app can make consumers' shopping experience more convenient and enjoyable, encouraging them to make larger purchases and spend more time shopping, which leads to more buying.

Technological developments have brought significant changes in various aspects, including shopping. Technologies include payment methods, tools to help customers, and even internet shopping that offers a shopping experience through smartphone apps [2]. In particular, retail apps can foster meaningful relationships between customers and retailers by providing various benefits to customers [3]. Technological development has been accompanied by

***Corresponding author: Esti Dwi Rinawiyanti (estidwi@staff.ubaya.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i1.453

adopting artificial intelligence (AI) and the increasing number of internet users in Indonesia [4]. At the beginning of 2024, there were 185.3 internet users and 139 million social media users in Indonesia [5]. AI can significantly improve the online grocery shopping experience for customers and retailers by offering personalized product recommendations based on customer behavior and interests and assisting customers in finding new items they would find interesting [6].

Regarding apps, to be effective and in line with user needs, serious attention to user interface (UI) and user experience (UX) is essential [7] because an attractive appearance of apps can have a significant psychological impact on users [8]. Thus, the main focus is designing optimal UI and UX that are comfortable and easy for users to understand [9], [10]. UI relates to the appearance design of the app, while UX focuses on design that aims to increase user satisfaction when using the app [11]. Regarding the item search app, AI can help consumers find the location of goods and direct them to the goods they are looking for via the fastest route. AI display design is equipped with UI/UX to make the app's appearance more attractive [12]. Notably, the supermarket digital app should be able to operate intelligently, offering optimal routes based on user behavior and preferences, whether the customer seeks a swift purchase or desires further product exploration. UI/UX design in AI-based digital apps will provide significant efficacy in its app. Usability is a quality indicator determining whether a product's UI is appropriate for its target users and can meet the fundamental needs of its users [13].

This study uses mixed methods, including qualitative and quantitative analysis, to determine the usability of supermarket goods search apps. The usability testing is conducted quantitatively using Maze software and a questionnaire, focusing on five variables: learnability, efficiency, memorability, accuracy and errors, and satisfaction. A qualitative analysis was also conducted to obtain the users' overall opinions and complete the quantitative data.

The remaining sections of this paper are organized as follows: Section 2 explains the theoretical underpinnings of AI, UI/UX, and usability testing, followed by Section 3, which describes the research methods. Section 4 presents the results of qualitative and quantitative usability testing. Section 5 discusses the findings, while Section 6 presents the study's limitations, conclusions, and recommendations for future research.

2. Literature Review

2.1. Artificial Intelligence

Artificial Intelligence (AI) is used to imitate the intelligence possessed by living creatures and inanimate objects to carry out commands. AI is a framework or system in an app designed to provide convenience and help solve problems [14]. AI has several benefits, including greater permanence, ease of duplication and dissemination, faster operation, and better outcomes. AI refers to developing systems that can perform tasks requiring human intelligence with two mains types [15]. First, narrow AI is a type of AI designed to perform a specific task or a group of closely related tasks. This AI is limited in scope and cannot generalize its intelligence to tasks outside its defined domain, such as image recognition software, translation tools, and virtual personal assistants, such as Alexa or Siri. Second, general AI, also known as strong AI, is a subset of AI that exhibits intelligence comparable to human intelligence in various tasks and domains. This AI can understand, learn, and apply knowledge in a way that is indistinguishable from human intelligence.

The fourth retail industry development, Retail 4.0, uses 4.0 technologies to satisfy the customers' needs, including AI, Internet of Things (IoT), cloud computing, big data analytics, and augmented reality [16]. The combination of innovation, technology, and people have significantly improved data analytics and production technologies [17]. Retail 4.0 technologies can help businesses stay competitive by improving customer experience and performance and offering customers a better shopping experience by integrating technology like AI and IoT into physical stores [18], [19]. AI-driven automation in retail will reveal a plethora of potential to enhance operational efficiencies, customer experience quality, and service levels through the intelligent utilization of valuable data generated at multiple points [20].

2.2. UI and UX

Designing UI/UX is crucial, especially in digital app design, because it provides the user's initial impression [7]. UI refers to interface design that prioritizes aesthetics, including selecting appropriate colors, layout, and logo design, as well as other measures to improve the visual appeal of a website or app, thereby enhancing user comfort and engagement [12], [21]. In addition to being visually appealing, the UI should be user-friendly or straightforward for users to use, enables users to carry out specific tasks more efficiently, such as navigating, searching for information, and interacting with elements in the app, improves UX, and provides a favorable impression of the product or digital platform [22], [23]. Thus, UI designs significantly impact the app's usability. Misunderstandings may result from poorly designed AI [24].

Meanwhile, UX can be defined as the degree to which a product, service, or system satisfies its target users through factors such as its appearance, accessibility, the efficacy of its UI, and support from previous experiences. UX combines feelings, beliefs, preferences, perceptions, physical and psychological reactions, behavior, and achievements before, during, and after using the product [24].

2.3. Usability Testing

Usability is the extent to which specific users can use a product to achieve certain goals through effectiveness, efficiency, and satisfaction, viewed from a particular use context [24]. Usability is a quality attribute that describes or measures how easy it is to use an interface. In the context of the design process, usability can also refer to a strategy that aims to make the product more user-friendly [25]. Usability includes the system interface and the interaction between the system and the user. The need for usability arises with the continuous increase and change in the user profiles of the products that develop with technology because it is an essential factor in designing interactive products and in developing and evaluating the quality of a system or app seen from the UX [27], [28], [29].

Usability testing is an evaluation process to measure how users can use a product or app effectively, efficiently, and satisfactorily. The purpose of usability testing is to evaluate UX directly, identify problems or difficulties that users may face, and provide recommendations for improvements to enhance the product or app [30], [31]. Several studies identified five key aspects of usability testing: learnability, efficiency, memorability, errors, and satisfaction [32], [33], [34], [35]. Learnability measures how users can effectively execute tasks upon initial interaction with the interface. Efficiency is assessed based on task completion speed, cost-effectiveness, and accuracy, and associated with performance. Memorability refers to the capacity to become accustomed to an app after a period of discontinuation, followed by a return to the same app. Errors are identified by user mistakes occurring during system utilization. A digital app must possess functional features to guarantee customer satisfaction, as users expect error-free experiences when utilizing apps [27] that relate to the degree of happiness experienced by users when using the system. Effectiveness pertains to how a system or app achieves its intended purpose.

Usability testing includes quantitative and qualitative measurements [28]. Quantitative usability testing can be done through scenario testing, a great approach to learning how a product performs in real-world scenarios. In this method, participants are tested in various tasks representing daily activities. Then, the results are analyzed based on the task completed, time per completed task, error rate, and number of clicks to evaluate the extent to which the product design meets customer needs and expectations [33]. Quantitative usability testing can also employ the questionnaire [32], [36] to measure the five aspects of usability testing: learnability, efficiency, memorability, errors, and satisfaction [25]. Meanwhile, qualitative testing can be done through observation and direct interaction (interview) with participants (also called respondents) [31], [36].

3. Methodology

Usability testing in this study was conducted using a mixed method approach (figure 1) that combines complementary quantitative and qualitative analyses. The quantitative analysis was carried out by first establishing seven tasks that needed to be completed using Maze [37], followed by questionnaires, which include five key aspects of usability testing: learnability, efficiency, accuracy, memorability, and satisfaction [25], [32]. The results of the quantitative analysis will be supported by the qualitative analysis that summarizes participants' opinions regarding

 ISSN 2723-6471

the app, including problems, expectations, and suggestions for the future. Table 1 reveals the metrics applied in mixed-method usability testing in this study.

In this study, Figma was the chosen design program for the app prototype [38]. Then, it was tested through quantitative and qualitative usability testing to ensure it meets the user's needs. Moderated remote testing is used in this study, which refers to the presence of a facilitator during the test instructing the participant or asking them questions. It represented in-person testing where facilitators and participants occur in the exact location [37], [39]. During testing, the facilitators may use web-based software to instruct participants, track their progress as they complete tasks, and even pose questions [40]. The facilitators engaged participants in functions as needed and conducted physical observations of their interactions with the product [39].

Maze, a powerful web app tool explicitly designed for UX research and usability testing, was employed in this study because of its compatibility with a wide range of design tools, including Figma. As one of the top 15 best usability testing tools and software in 2024, Maze can simplify users' recruitment, streamline the usability testing process, and provide a window into how users experience the prototype. Maze facilitates the direct assessment of an app's usability by producing a comprehensive usability test report, including key metrics such as completion rates, misclick rates, and time spent. The report can inform enhancements to the app design [37].

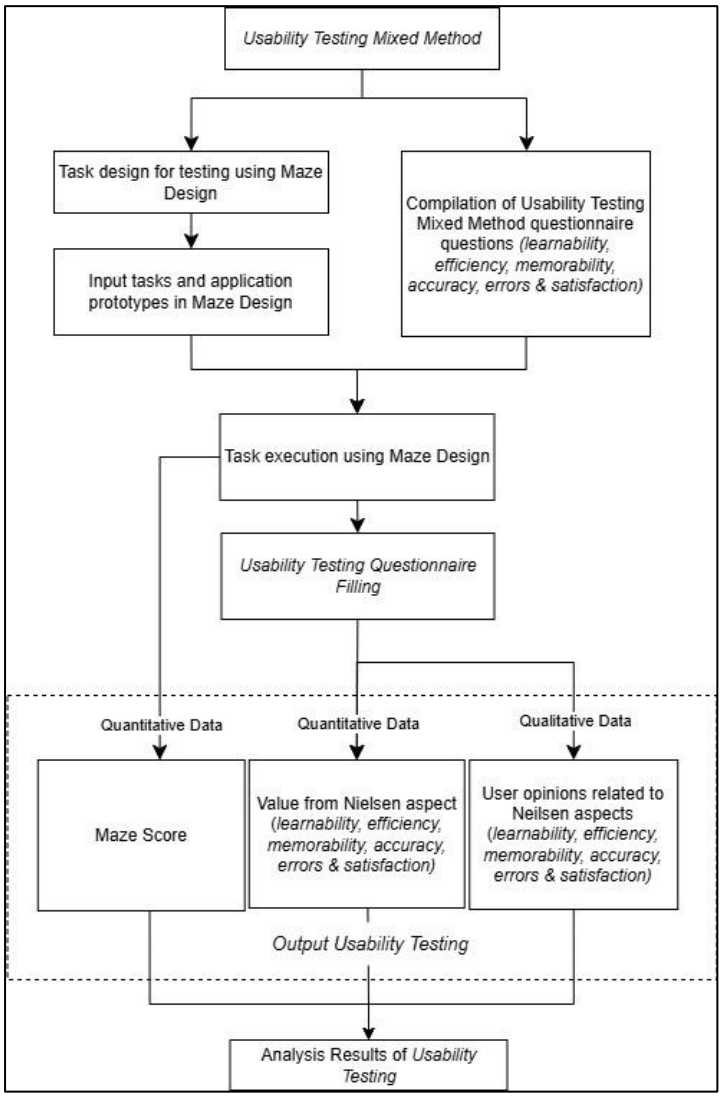

Figure 1. Mixed Method Usability Testing

The selection of participants (employed a purposive sampling technique, in which respondents were selected based on subjective personal judgment and convenience [40], such as conformity with the sample criteria. The respondents

 ISSN 2723-6471

or participants in this study should meet specific criteria, such as frequent shopping at supermarkets, online access, smartphone and internet connectivity, and online shopping.

According to data from the Central Statistics Agency, most Indonesians aged 25 and over (58.63%) have accessed the Internet in the last three months for various purposes, including work, shopping, entertainment, and communication. Meanwhile, Indonesians between the ages of 19 and 24 are in second place, with 14.69% having used the Internet in the previous three months. While the app's design aims to be user-friendly for youth and older people, ten chosen participants can represent these groups of Indonesian people aged 19 and over.

Usability tests typically consist of 6–12 participants split into 2-3 subgroups. Specifically, three to four persons from each category are recommended when assessing two user groups, and three people from each category are required when testing three or more user groups [25], [41]. Previous studies have also employed usability testing with 5-10 participants [42], [43], [44]. These findings support that ten participants were adequate in this study.

Table 1. Mixed Method Usability Testing Assessment Metrics

Aspect	Quantitative Assessment		Qualitative Assessment
	Usability Questionnaire	Maze	
Learnability	Percentage of user ease in understanding the given tasks and app system	-	Opinions regarding the level of ease in understanding the given tasks and app system
Efficiency	Percentage of user convenience in accessing menus and obtaining information in the app easily and quickly	Time required to complete each task	Opinions regarding the level of ease of completing the tasks given
Accuracy	Percentage of information accuracy in completing tasks and the role of functions in answering user needs in shopping.	Percentage of accuracy through the type of Task completion for each respondent (direct, indirect, mission unfinished)	Opinions on things that can cause users to experience errors when interacting with the app system
Errors	-	Percentage of user click errors when using the app	Analysis of the causes of click errors made by users
Memorability	Percentage of ease in remembering the location and use of features and in using the app at any time	-	Opinions regarding things that can be improved to improve the ability to remember features and the usefulness of features in the app
Satisfaction	Percentage of users who feel happy and satisfied with the appearance of the app	-	Opinions regarding things that need to be done to improve user satisfaction when using the app

4. Results and Discussion

The app prototype was developed utilizing Figma [38], a website that provides app design tools. Five principles—consistency, accessibility, simplicity, flexibility, and feedback—were applied. These principles are designed to enhance app intuitiveness and usability and facilitate respondents' achieving their objectives. Figure 2 presents an overview of the app prototype.

 ISSN 2723-6471

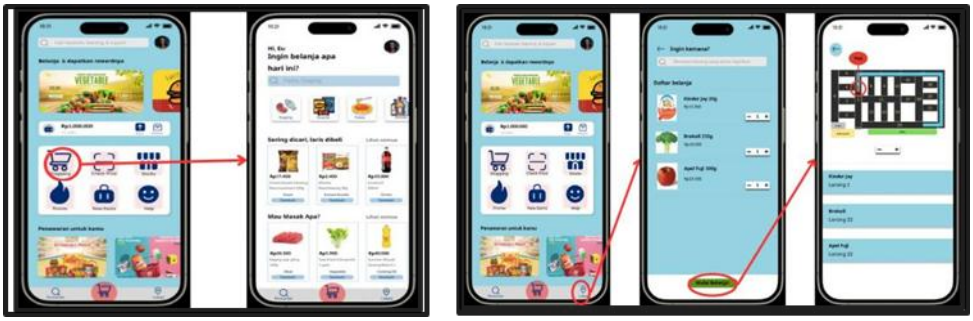

Figure 2. Prototype of Supermarket's Goods Search App

4.1. Quantitative Analysis Through Task Scenarios Using Maze Score

In this evaluation, the task scenarios were assigned during usability testing to assess to what extent the participants will perform when utilizing apps in their daily lives. They would complete tasks as they would in real life, regardless of whether they input actual data or click on predetermined areas (hot spots) by the predetermined scenario, even though the system utilized in this test is a simulation. The app's design will be evaluated by participants completing the subsequent tasks, as presented in table 2.

Table 2. Results of the Usability Testing Using Maze

No.	Purpose	Task	Usability Score (%)	Success Rate (%)	Misclick Rate (%)	Duration (seconds)
1.	Account registration	Register or login as a user	67	100	61.8	29.6
2.	Search for a specific product	Using the shopping and product search functions	50	90	46.4	40.2
3.	Search for the location of a particular product	Using location features	42	70	22.8	37.2
4.	Get promotional information	Looking for promotions or new product items	64	80	38.3	17.7
5.	Check product prices	Use the check-price feature	44	80	41	43.6
6.	Seek help from staff	Using the Help feature	95	100	9.1	3.3
7.	Know the availability of specific items	Using the Stock feature	53	80	38.7	25.6
Average			**59**	**85.7**	**36.9**	**28.2**

Scenario testing would measure four parameters: task completed, time per completed task, error rate, and number of clicks [36]. Task completion time is the time it takes a participant to complete a particular task in an app or system. It measures how long it takes participants to accomplish a specific task, which aids in evaluating how effective a UI/UX is [45]. The task completed reflects a success rate that includes [46], [47]: (i) successful: the task can be finished successfully, and there were fewer or the same number of processing step mistakes as the number of processing steps given; (ii) partially successful: tasks are those that have been finished successfully but have more errors in their processing steps than the number of predetermined processing steps; and (iii) failed: the task cannot be finished appropriately. Duration refers to time per completed task; time spent on participant tasks recorded from the moment the participant starts the assignment until they stop working on it. Misclick rate evaluates the error rate and number of clicks by calculating an average number of incorrect clicks for a specific path and screen [37].

The results of usability testing using Maze are presented in table 2. The usability score for task 1 (Register or login as a user) is 67% and a success rate of 100%. This result shows that all participants completed this task without difficulty, indicating that the steps required to register and log in are simple for users to understand and follow. The misclick rate for task 1 is 61.8%, implying that users need help with which part to click. This result could indicate

 ISSN 2723-6471

that interface elements need to be clarified or capture the user's attention optimally, resulting in frequent incorrect clicks and an average time of 29.6 seconds to complete the registration and login process. This is a relatively short duration, but it could be shortened.

In task 2 (Using the shopping and product search functions), the usability score is 50%, with a 90% success rate and 46.4% misclick rate. This result indicates that while some respondents completed the task, clicks were frequently made in unexpected locations, resulting in an average completion time of 40.2 seconds. This duration is unusually long, likely due to the respondent's bias in investigating the prototype display.

Meanwhile, the usability score for task 3 (Using location features) is 42%, with a success rate of 70%, a misclick rate of 22.8%, and a drop of 30%. This result means that some participants completed tasks differently than the flow. This result could be because task 3 has a relatively long page flow, and participants are interested in exploring the app prototype. However, not all buttons can be clicked, affecting the misclick rate and the usability score.

The usability score for task 4 (Looking for promotions or new product items) is 64%, with a success rate of 80% and a misclick rate of 38.3%. This indicates that the participants completed the task, which took an average of 17.7 seconds. However, not all of them could finish the expected efficient flow. This was because the participants were exploring the display, which resulted in click errors that impacted the usability score.

Moreover, the usability score for task 5 (Use the check-price feature) is 44%, and the average time it takes to complete the task is 43.6 seconds. The success rate for this task is 80%. This result demonstrates that it took the participants a considerable amount of time to comprehend how to access the page where they could check the price, and they frequently made clicking errors, which affected the usability score. It should be noted that many participants could still finish the task.

Completing task 6 (Using the Help feature) yields a usability score of 95%, a success rate of 100%, and a click error rate of 9.1%. Given that the average duration of this task is relatively short, this result implies that it is easy for participants to follow so that they can successfully carry out the task quickly and with minimal click errors.

Regarding task 7 completion (Using the Stock feature), a usability score of 53% was generated, which indicates that it was pretty low because it had a misclick rate of 38.7%. With a high success rate of 80% and an average duration of 25.6 seconds, this task was successfully finished from start to finish.

Table 2 and figure 3 show that the average success rate of the 7-task completion is 85.7%, with an overall Maze usability score of 59 %. The thresholds used to classify the Maze usability score are high (80-90), medium (50-80), and low (0-50). Thus, the final Maze usability score indicates medium usability. The result reveals several enhancements required for the prototype.

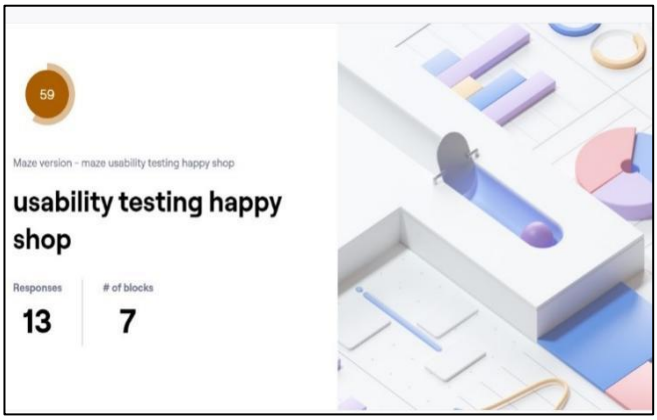

Figure 3. Final MAZE Score

4.2. Quantitative Analysis Using Five Aspects of Usability Testing

Upon completing seven tasks using Maze, participants (respondents) were required to complete a questionnaire to evaluate their opinions and assessments of the tested app. Five aspects of Nielsen were involved in the questionnaire: learnability, efficiency, accuracy and errors, memorability, and satisfaction since these five aspects were applied in

several previous studies [25], [35], [42], [48], [49]. Besides, numerous established software usability models and standards frequently discuss learnability, efficiency, effectiveness, and satisfaction [50].

The questionnaire had 15 questions utilizing a scale from 1 (strongly disagree) to 5 (strongly agree) to measure participant responses. Table 3 presents the results of five aspects of measurement. The learnability aspect, which indicates how easily new users can understand and start using the system effectively, gets an average score of 85%. The high percentage suggests that the app is easy for users to comprehend. Meanwhile, the efficiency aspect, which measures how quickly users can complete tasks after learning the system, obtains an average score of 80%. This result indicates that respondents can understand the system quickly and efficiently.

Table 3. Measurement of Five Aspects of Usability Testing

Aspect	Question	Average
Learnability	High level of ease in understanding the tasks given. High level of ease in understanding the app system.	85
Efficiency	Users can access the menu quickly. Users can get information in the app easily and quickly.	80
Accuracy and Errors	Task completion is accompanied by inaccurate information, which confuses completion. The tasks given are appropriate and can meet user needs when app shopping.	60
Memorability	Users easily remember the app's features' location and functionality. The app is user-friendly and accessible.	85
Satisfaction	Users feel happy and satisfied with the app's appearance.	50
	Average Usability Score	**72**

In addition to the results of usability testing obtained from the questionnaire, the duration that the respondents spent on each task obtained from Maze can also serve as a reference for subsequent improvements related to efficiency. The average time each respondent needs to pay to finish the seven tasks given is 30,557 seconds, with a few data variations. During the subsequent improvement, it is anticipated that this will be decreased, allowing for the average duration to be shortened and the variations in the data to be more consistent.

As shown in table 3, the average accuracy value is 60%. Maze results supported this finding by providing information on respondents who could complete the task with the expected flow successfully, those who could not complete the task, and those who could not complete the task with the expected flow. With the help of this information, it is possible to calculate the accuracy percentage for each respondent and app prototype. The Maze website generated a value of 70.714%, which implies that participants still do not fully complete the task with the expected flow, indicating a need for improvement.

The error aspect is always paired with accuracy. Errors can be seen from the mistakes made by participants in carrying out the tasks given. The misclick rate data can represent this result from the Maze website (see table 2). The misclick rate has a negative correlation; as the misclick rate increases, the frequency of user errors escalates, adversely impacting the usability score. A lower misclick rate indicates users can navigate the app seamlessly with few click failures. Table 2 indicates that the average error or misclick rate is 36.9%, suggesting several participants frequently commit clicking errors throughout the task completion. Clicking errors in usability testing may arise from the placement of unfamiliar features and users' intention to explore. However, not all prototype buttons are operational, contributing to click errors as identified by the Maze Website. The average misclick rate is relatively low, although the data has significant variability.

In terms of memorability, it measures users' ability to recall system usage after a period of inactivity. This aspect got an average score of 85%. Regarding satisfaction, which measures participants' overall satisfaction with the tested app, the result shows that 50% of participants are satisfied. This value indicates that there is a need for improvement, particularly in terms of design, to increase the level of satisfaction experienced by users. Overall, usability testing using questionnaires had an average score of 72%, implying a medium usability score.

 ISSN 2723-6471

4.3. Qualitative Analysis of Usability Testing

Qualitative usability testing was conducted via observation and interviews. Observations were carried out during the task scenarios. Facilitators observed participants to find out what happened during the task scenarios. Subsequently, interviews were employed following the completion of the task to ascertain the participants' behavioral experiences during the assessment [36]. Qualitative usability testing would be used as a reference to determine strengths/advantages, weaknesses, suggestions, or recommendations for improvements to the prototype [51]. Table 4 shows the result of the qualitative analysis.

Table 4. Qualitative Assessment Results

Aspect	Feedback	Recommendation
Learnability	The display is less attractive and rigid.	Increase icon size, color, and font.
	Features are easy to understand but similar to general apps.	Use more intuitive icons.
	Some users feel confused by feature placement and page-switching.	Fix confusing feature placement.
	Location and cart features are confusing.	Add a search bar to the location feature.
Efficiency	Some tasks are easy to understand and quick to complete.	Fix lag issues that interfere with smooth operation.
	The interface helps locate features but suffers from occasional lags.	Move the location icon to a strategic position.
	The location feature is poorly positioned.	
Accuracy	The app lacks usage mix clarity.	Add app usage guides.
	Icons are too big, and description fonts are too small.	Organize feature layouts strategically.
	Some features are not placed strategically.	
Memorability	The app design is less eye-catching, making features harder to remember.	Improve visual design with striking colors.
	Feature icons are less prominent and hard to differentiate.	Enhance color and shape differentiation.
	Icons used are not familiar to users.	
Satisfaction	The app is helpful but could use additional features.	Add a voice notification feature for interactive experience.
	Navigation is difficult for some tasks.	Display product location in item descriptions.
	Colors lack contrast and striking visual appeal.	Include fresh product arrival date information.

Based on the results of qualitative analysis in usability testing, participants faced several critical issues related to the placement of features and the app's appearance. Participants generally felt they could complete the assigned tasks quickly with the straightforward instructions provided. Nevertheless, some participants needed help finding certain features, especially those that are rarely used or have poor navigation within the app. Most participants found this app relatively easy to use and understand. Still, some have complained about the less appealing visual design and the size of the icons or text too small, which makes UX less comfortable, especially for older participants.

Other obstacles included unclear feature placement and less eye-catching. New participants, in particular, needed more clarity when trying to understand the flow of the app with adequate initial guidance. Nonetheless, the overall satisfaction level is quite good; however, many participants hoped for improvements in visuals and navigation and additional features that could enrich their shopping experience, such as a favorite feature and more detailed product descriptions. Participants hope the next prototype will have a more appealing design, with more transparent and more distinguishable icons, colors, and layouts. Additionally, the inclusion of an initial user guide and an increase in font and icon sizes are also expected to enhance the app's usability, especially for participants with visual impairments or older individuals.

5. Conclusion

This study aims to design a supermarket goods search app using design thinking and a human-centered approach. By focusing on the UI/UX, the app enables supermarket consumers to find the location of goods and improves their shopping experience. The prototype was created using Figma by designing an attractive UI/UX. First, usability testing was conducted with task scenarios using Maze, which resulted in an average Maze score of 59%, suggesting a medium usability score. Second, usability testing was carried out using questionnaires with positive results, including high scores on learnability (85%), efficiency (80%), and memorability (85%) but medium scores on accuracy and errors (60%) and satisfaction (50%). The average score is 72%, indicating a medium usability score. These results suggest that although the app can meet user needs and increase the shopping experience, it needs to be improved. Third, usability testing was conducted qualitatively through observations and interviews. From qualitative analysis, participants have suggested some expectations for future improvements in this digital app to enhance customers' shopping experience regarding layout and ease of use, as well as additional features like a favorite section and more product descriptions.

There are several limitations to this study. First, this study only created a prototype design for the app's features and appearance using Figma. Further research can complete the app with programming coding or even to systems. Second, usability testing in this study included only ten participants. Future research may incorporate a larger sample size within each category (students, employees, housemakers) for more complete findings. Third, this study was restricted to one supermarket. Further research can be extended to other supermarkets or objects, such as public facilities ranging from malls to airports, where it is difficult for consumers to find and remember the desired location.

6. Declarations

6.1. Author Contributions

Conceptualization: E.D.R., R.M.S., M.H., E.P.P., A.C.V.C.; Methodology: R.M.S., M.H.; Software: E.D.R.; Validation: E.D.R., R.M.S., and A.C.V.C.; Formal Analysis: R.M.S., M.H., E.P.P, and A.C.V.C.; Investigation: E.D.R., A.C.V.C.; Resources: R.M.S., M.H., and E.P.P.; Data Curation: R.M.S., E.P.P., A.C.V.C.; Writing—Original Draft Preparation: E.D.R., R.M.S., and A.C.V.C.; Writing—Review and Editing: R.M.S., E.D.R., and A.C.V.C.; Visualization: R.M.S. and A.C.V.C. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors would like to thank Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi (DIKTI) for their support in funding this study.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Nurhayati-Wolff, "Supermarket sales in Indonesia 2017-2022." statista.com. (accessed Oct. 15, 2024).

[2] H. Y. Purwantono, A. A. S. Gunawan, H. Tolle, M. Attamimi, and W. Budiharto, "A literature review: Feasibility Study of technology to improve shopping experience," *Procedia Computer Science*, vol. 179, no. 2, pp. 468-479, 2021.

[3] A. K. A. Farha, K. E. Hedhli, I. Alnawas, H. Zourrig, and I. Becheur, "Drivers and outcomes of a shopper-retailer's app relationship," *Journal of Retailing and Consumer Services*, vol. 81, no. 2, pp. 1-18, 2024.

[4] O. Tyrväinen, H. Karjaluoto, "Online grocery shopping before and during the COVID-19 pandemic: A meta-analytical review," *Telematics and Informatics*, vol. 71, no. 2, pp. 1-13, July 2022.

[5] World Bank Group, "Individuals using the Internet (% of population)–Indonesia." worldbank.og. https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=ID (accessed Nov. 25, 2024).

[6] D. Aggarwal, "Exploring the Role of AI for Online Grocery Shopping through Enhancing Personalized Recommendations and Customer Segmentation," *Technoarete Transactions on Advances in Computer Apps (TTACA)*, vol. 2, no. 3, pp. 1-4, Sep 2023.

[7] R. Gunawan, G. Anthony, Vendly and M. S. Anggreainy, "The Effect of Design User Interface (UI) E-Commerce on User Experience (UX)," *Proceeding of 2021 6th International Conference on New Media Studies (CONMEDIA)*, vol. 2, no. 3, pp. 95-98, 2021, doi: 10.1109/CONMEDIA53104.2021.9617199.

[8] B. Upasna, C. Klarissa, N. Tillmann, "Understanding the impact of perceived visual aesthetics on user evaluations: An emotional perspective," *Information and Management*, vol. 56, no. 1, pp. 85-93, 2019, https://doi.org/10.1016/j.im.2018.07.00.3.

[9] I. Darmawan, M. S. Anwar, A. Rahmatulloh, H. Sulastri, "Design Thinking Approach for User Interface Design and User Experience on Campus Academic Information Systems," *International Journal on Informatics Visualization*, vol. 6, no. 2, pp. 327-334, 2022.

[10] E. F. Firdani, I. F. Hanif, Muhammad Febriansyah, and M. Y. Widsono, "Implementation of Website-Based UI UX Using the Design Thinking Method: Case Study at PT Jesinra Makmur Group," *Journal of Advances in Information and Industrial Technology*, vol. 6, no. 1, pp. 73–82, Aug. 2024.

[11] R. Maringka and C. Lumingkewas, "Designing User Interface (UI) And User Experience (UX) of a Sport Space Rental Application using Design Thinking Method," *CogITo Smart Journal*, vol. 10, no. 1, pp. 613–624, 2024, doi: 10.31154/cogito.v10i1.692.613-624.

[12] R. A. Bertão, J. Joo, "Artificial intelligence in UX/UI design: a survey on current adoption and [future] practices," *Proceeding of the 14th International Conference of the European Academy of Design (EAD), Safe Harbours for Design Research 2021*, vol. 9, no. 5, pp. 404-413, Dec. 2021, doi: 10.5151/ead2021-123.

[13] N. Sun and C. Sun, "On the Evaluation of the Usability of E-commerce UI Design," *in Proceedings of the 2nd International Conference on Education, Language and Art (ICELA 2022), Paris: Atlantis Press SARL*, vol. 2, no. 1, pp. 1000–1011, 2023, 10.2991/978-2-38476-004-6_120.

[14] Å. Stige, E. D. Zamani, P. Mikalef, and Y. Zhu, "Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future research agenda," *Information Technology and People*, vol. 37 no. 6, pp. 2324-2352, 2024.

[15] M. Damar, A. Özen, Ü. E. Çakmak, and E. Özoğuz, "Super AI, Generative AI, Narrow AI and Chatbots: An Assessment of Artificial Intelligence Technologies for The Public Sector and Public Administration," *Journal of AI*, vol. 8, no. 1, pp. 83-106, 2024.

[16] P. Sakrabani, A. P. Teoh, and A. Amran, "Strategic impact of retail 4.0 on retailers' performance in Malaysia," *Strategic Direction*, vol. 35 no. 11, pp. 1-3, 2019.

[17] S. S. Kamble, A. Gunasekaran, A. Ghadge, and R. Raut, "A performance measurement system for industry 4.0 enabled smart manufacturing system in SMMEs-A review and empirical investigation," *International Journal of Production Economics*, vol. 229, no. 107853, pp. 1-15, 2020, doi: 10.1016/j.ijpe.2020.107853.

[18] L. L. Har, U. K. Rashid, L. T. Chuan, S. C. Sen, and L. Y. Xia, "Revolution of Retail Industry: From Perspective of Retail 1.0 to 4.0," *Procedia Computer Science*, vol. 200, no. 4, pp. 1615-1625, 2022, doi: 10.1016/j.procs.2022.01.362.

[19] P. Sakrabani and A.P. Teoh, "Retail 4.0 adoption and firm performance among Malaysian retailers: the role of enterprise risk management as moderator," *International Journal of Retail and Distribution Management*, vol. 49 no. 3, pp. 359-376, 2021.

[20] K. Kapate, "Retail Intelligent Automation: An overview," *International Journal of Innovative Science and Research Technology (IJISRT)*, vol. 4, no. 7, pp. 887-891, 27 July 2019.

[21] A. P. Priyadarshini, "The Impact of User Interface Design on User Engagement," *International Journal of Engineering Research and Technology (IJERT)*, vol. 13, no. 03, pp. 1-11, March 2024.

[22] J. Wang, Z. Xu, X. Wang and J. Lu, "A Comparative Research on Usability and User Experience of User Interface Design Software," *International Journal of Advanced Computer Science and Applications,* vol. 13, no. 8, pp. 21-29, 2022.

[23] T. Tepe, "Students' experiences and usability evaluation in interactive digital interface development process," *Journal of Computer and Education Research,* vol. 10, no. 20, pp. 434-451, 2022.

[24] A. Kamil, H. Hassan, H. Hassan, R. A. Jamaluddin, and J. M. Polohan Nadan, "UI and UX Redesign for COVID-19 Response Mobile Application Using Usability ISO Model and Testing Tools," *Open International Journal of Informatics (OIJI),* vol. 11, no. 2, pp. 210–223, Dec. 2023.

[25] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *Journal of Interaction Science,* vol. 1, no. 1, pp. 1-16, 2013.

[26] Y. Rogers, Sharp, and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 3rd ed., NY, USA: John Wiley and Son, Ltd., 2002.

[27] H. Anam, M. Sadiq, and H. Jamil, "Development of System UsabilityScale for the Urdu Language," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 18 no. 6, pp.73-78, 2020.

[28] M. F. Alwashmi, J. Hawboldt, E. Davis, and M. D. Fetters, M. D., "The Iterative Convergent Design for Mobile Health Usability Testing: Mixed Methods Approach", *Journal of Medical Internet Research mHealth uHealth,* vol. 7 no. 4, pp. 1-15, 2019.

[29] B. Maqbool, and S. Herold, "Potential effectiveness and efficiency issues in usability evaluation within digital health: A systematic literature review," *The Journal of Systems and Software*, vol. 208, no. November 2023, p. 111881, 2024, doi: 10.1016/j.jss.2023.111881.

[30] J. Dumas and J. C. Redish, A practical guide to usability testing, OR, USA: Intellect, 1999.

[31] A. P. Nugraha, Rolando, M. A. Puspasari, and D. H. Syaifullah, "Usability Evaluation for User Interface Redesign of Financial Technology Application," *in IOP Conference Series: Materials Science and Engineering, Dec. 2019*, vol. 505, no. 1, pp. 12101-12109, doi: 10.1088/1757-899X/505/1/012101.

[32] A. M. H. Kilis, J. Mamonto, H. Legesan, and Z. Sagai, "Usability Evaluation of The Android Operating System Using Use Questionnaire", *International Journal of Information Technology and Education (IJITE)*, vol. 1 no. 2, pp. 16- 23, March 2022.

[33] A. Madan and S. K. Dubey, "Usability evaluation methods: a literature review", *International Journal of Engineering Science and Technology*, vol. 4, no. 2, pp. 590-599, Feb. 2012, https://doi.org/10.1002/9781119636113.ch38.

[34] S. Möller, Usability Engineering, In: Quality Engineering. Berlin, Germany: Springer, 2023.

[35] N. B. Puspitasari, Y. Aulia and Z. F. Rosyada, "Usability evaluation of online transportation using Nielsen model", *World Journal of Advanced Engineering Technology and Sciences*, vol. 10, no. 02, pp. 1-9, 2023.

[36] J. R. Lewis, and J. Sauro, Usability and User Experience: Design and Evaluation, In Handbook of Human Factors and Ergonomics (eds G. Salvendy and W. Karwowski), NJ, USA: Wiley, 2021, doi: 10.1002/9781119636113.ch38.

[37] A. R. Najaf, R. Permatasari, and C. Sakti, "Evaluating the Usability of a Book Stock and Sales Recording Android App: System Usability Scale (SUS), Heuristic Evaluation, and Maze Tool Insights", *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1774-1789, Sep. 2024, doi: 10.51519/journalisi.v6i3.829.

[38] Z. E. Putra, H. Ajie, and I. A. Safitri, "Designing A User Interface and User Experience from Piring Makanku Application by Using Figma Application for Teens." *IJISTECH (International Journal of Information System and Technology),* vol. 5 no. 3, pp. 308-315, 2021.

[39] L. Riani, A. Yuaziva, M. R. Rusafni, A. Fami, and I. R. G. Barus, "Application of Moderated Remote Usability Testing Method on Agrovest Website", *EduTech: Jurnal Teknologi Pendidikan,* vol. 23 no. 2, pp. 177-188, 2024.

[40] S. D'Alessandro, H. Winzar, B. Lowe, and W. Zikmund, Marketing Research, 5th Edition; Publisher, Cengage AU, 2020.

[41] C. W. Turner, J. R. Lewis, and J. Nielsen, Determining usability test sample size, In: W. Karwowski (Ed.), International Encyclopedia of Ergonomics and Human Factors, 2nd Edition, vol. 4, Boca Raton, FL, USA: CRC Press, pp. 3084–3088, 2006.

[42] P. Weichbroth, "Usability Testing of Mobile Applications: A Methodological Framework," *Applied Sciences*, vol. 14, no. 5, pp. 1-27, 2024, doi: 10.3390/app14051792.

[43] M. G. Sidiq, R. G. Guntara, and O. Herdiana, "The Evaluation of User Experience UPI Digital Business Website with Usability Testing Method and System Usability Scale", *Indonesian Journal of Digital Business*, vol. 4 no. 1, pp. 1-10, 2024.

[44] G. J. P. Hakim, F. Saputra, M. N. Ardhani, A. Fami, and I. R. G. Barus, "V-Care Application Design Using Design Thinking Method", *Journal of Informatics and Communication Technology (JICT)*, vol. 6, no. 1, pp. 185-197, Oct. 2024, doi: 10.52661/j_ict.v6i1.291.

[45] N. A. Nik Ahmad, N. I. M. Hamid, and A. Mohd Lokman, "Performing Usability Evaluation on Multi-Platform Based Application for Efficiency, Effectiveness, and Satisfaction Enhancement", *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 10, pp. 103–117, May 2021.

[46] A. Liu, X. Song, and P. Hansen, "Characterising users' task completion process in learning-related tasks: A search pace model", *Journal of Information Science*, vol. 49, no. 6, pp. 1462-1480, 2023.

[47] P. Weichbroth, "Usability of Mobile Applications: A Consolidated Model," *IEEE Access*, vol. 12, no. April, pp. 62378–62391, 2024, doi: 10.1109/ACCESS.2024.3395528.

[48] H. Christanto, S. Sutresno, and V. Suandi, "Usability Analysis of Atma Jaya Catholic University E-Learning Based on Human Computer Interaction", *Journal of Information Systems and Informatics*, vol. 5 no. 3, pp. 833-847, 2023.

[49] L. Faulkner, "Beyond the five-user assumption: benefits of increased sample sizes in usability testing", *Behavior Research Methods, Instruments, and Computers*, vol. 35 no. 3, pp. 379-383, Aug. 2003, doi: 10.3758/bf03195514.

[50] P. Weichbroth, "Usability of mobile applications: A systematic literature study," *IEEE Access*, vol. 8, no. March, pp. 55563–55577, 2020, doi: 10.1109/ACCESS.2020.2981892.

[51] P. O. Løvsletten, B. S. Hunskaar, A. F. Heen, G. Bekkering, L. V. Poel, D. Zeraatkar, M. Vermandere, B. Aertgeerts, N. Delvaux, F. Achille, J. W. Busse, T. Agoritsas, and P. O. Vandvik, "Physicians found an interactive tool displaying structured evidence summaries for multiple comparisons understandable and useful: a qualitative user testing study", *Journal of Clinical Epidemiology*, vol 172, no. 111399, pp. 1-10, Aug 2024, doi: 10.1016/j.jclinepi.2024.111399.