# Usability Evaluation of the Alodokter Application Using the Task-Based Usability Testing Method

**Achmad Dwiki Saputra**
Informatics Engineering Study Program STMIK IM
Jl. Belitung No. 7 Bandung
Email: dwkysap@gmail.com

## *ABSTRACT*

*Digital health applications such as Alodokter have become essential platforms for Indonesians to access healthcare services. Although mobile health adoption continues to rise, very few studies have empirically examined how design inefficiencies affect users' ability to navigate between medical and commercial service flows, which may reduce user trust and engagement in digital healthcare platforms. Research evaluating user friction during transitions between medical flows (such as Doctor Chat) and commercial flows (such as Aloshop) within an integrated ecosystem remains limited. This study applies the Task-Based Usability Testing (TBUT) method in the Indonesian digital health context to capture fundamental user interactions beyond perception-based evaluations. The objective is to systematically evaluate the usability level of the Alodokter application by involving real users performing representative tasks across four main flows: Doctor Chat, Appointment Booking, Article Search, and Medicine Purchase through Aloshop. The evaluation employed quantitative metrics (success rate and time on task) and qualitative observation using the Think Aloud Protocol. Results show that the Appointment Booking flow had the lowest usability (success score 1.76), while the Aloshop flow was less efficient, requiring 21 seconds longer than comparable applications. The main issues identified are the lack of insurance information and inefficient cart flow. This study is limited to task-based evaluation and does not include longitudinal analysis of user behavior. These findings provide empirical foundations for data-driven design improvements to enhance user experience and advance usability evaluation methods for hybrid healthcare-commerce platforms in Indonesia.*

*Keywords:* *Usability Testing, User Experience, Health Application, Alodokter, Task-Based Test*

## Introduction

The rapid advancement of digital technology has transformed the way people access healthcare services. *Mobile health* (mHealth) and telemedicine applications have become essential tools connecting patients and healthcare providers, particularly following the acceleration of digital adoption during the COVID-19 pandemic. In Indonesia, platforms like Alodokter play a central role, offering online consultations, appointment scheduling, and the purchase of pharmaceutical products. However, the quality of *the user experience* (UX) remains a key factor influencing user trust, convenience, and adoption rates, and users still encounter obstacles when navigating multiple service flows within a single application. [1]

While numerous studies have assessed *the usability of* digital health apps in Indonesia, very few have empirically examined how design inefficiencies impact users' ability to navigate between medical service flows (e.g., Doctor Chat, Appointment Booking) and commercial service flows (e.g., Aloshop) within a single, integrated ecosystem. Addressing this gap is crucial to understanding how hybrid healthcare-commercial platforms can deliver seamless, trustworthy, and user- centric experiences.

The concept of *usability,* as defined by ISO 9241-11, refers to the extent to which a specific user can use a product to achieve particular goals effectively, efficiently, and satisfactorily within a defined usage context. *Usability evaluation* is crucial in healthcare applications because interaction barriers can decrease trust and system adoption rates. [2]. Several studies have explored *usability* in the context of digital health. For example, Iskandarsyah et al. (2022) found that clear task flows and visual simplicity increased user satisfaction with the Jaga Sehat application, while [3] Emphasized the importance of integrating quantitative and qualitative feedback in improving telepharmacy design. International studies by [4] [5] and [6] Also highlighted efficiency, *learnability,* and interaction flow as determinants of user engagement in mHealth applications.

Despite these advances, most *usability studies* in Indonesia still rely on perception-based instruments such as *the System Usability Scale* (SUS), which are unable to capture real-world interaction behavior. [7]. Only a few studies have used task-based observation methods, which can uncover contextual interaction barriers. This study

provides a methodological contribution by applying *Task-Based Usability Testing* (TBUT) to analyze *cross-flow usability* in an integrated health-commercial platform—an approach rarely used in mHealth research in Indonesia.

While several usability studies have examined digital health applications in Indonesia, most have focused on general usability perceptions using standardized questionnaires or evaluated isolated features without considering cross-flow interactions. Previous studies on Alodokter and similar mHealth platforms primarily relied on perception-based measures and did not systematically analyze real user behavior across integrated medical and commercial service flows. This study introduces a novel contribution by applying Task-Based Usability Testing (TBUT) to evaluate multiple core user flows within a single hybrid mHealth ecosystem. By capturing task performance metrics and observed cognitive challenges during fundamental interactions, this research provides new empirical insights into cross-flow usability issues that have not been explicitly addressed in prior mHealth usability studies in the Indonesian context.

Despite the growing body of usability research on digital health applications in Indonesia, existing studies predominantly rely on perception-based evaluation methods and tend to assess single service features in isolation. To the best of the authors' knowledge, no prior study has systematically examined usability performance across multiple integrated medical and commercial flows within the Alodokter application using task-based behavioral data. This study addresses this gap by applying Task-Based Usability Testing (TBUT) to analyze fundamental user interactions across four core service flows within a hybrid mHealth ecosystem. By capturing both quantitative task performance and qualitative cognitive processes, this research offers a novel cross-flow usability perspective that extends beyond self-reported usability perceptions and provides new empirical insights into the design of an integrated healthcare-commerce platform in Indonesia.

TBUT evaluates *usability* through direct user engagement in specific, goal-oriented tasks [8] emphasizing observable performance measures such as effectiveness and efficiency while capturing subjective factors such as satisfaction [9]. This method is particularly well-suited for digital health systems because it uncovers practical *usability issues* often missed by perception-based approaches. [10].

Thus, this study aims to evaluate *usability* and identify key pain points in four primary Alodokter user flows—Doctor Chat, Appointment Booking, Article Search, and Aloshop—using the TBUT method. *Usability* is assessed based on three ISO 9241-11 dimensions: effectiveness (success rate), efficiency (task completion time), and satisfaction (usability *score)*. This study is expected to provide empirical insights into *usability performance* in hybrid healthcare-commercial systems and generate design recommendations that can be implemented to improve digital healthcare applications in Indonesia.

## Research Methods

This study uses *a mixed-methods approach,* integrating quantitative and qualitative techniques, to gain a comprehensive understanding of *the usability* of the Alodokter mobile health application. This study employed *Task-Based Usability Testing,* which evaluates user performance by having users complete representative tasks. This approach has been widely used and validated in healthcare contexts to uncover real-world interaction barriers. [11].

### Target / Research Subject

Participants were recruited purposively to ensure compliance with the study objectives. Ten individuals (four men and six women) aged 25–38 years residing in the Greater Jakarta area participated in this study. All participants were active or recent users of the Alodokter app. The sample size was determined based on [12] guidelines stating that 5–10 participants are sufficient for *formative usability testing*. Recent findings in health app usability research also support this guideline. [13]. The demographic characteristics of the participants are summarized in Table 1.

**Table 1**. Demographics of Research Participants

| Participant | Gender | Age | Familiarity with Alodokter |
|---|---|---|---|
| P1 | Woman | 28 | Active Users (≥ 1x/month) |
| P2 | Woman | 25 | Infrequent Users (A few times/year) |
| P3 | Man | 30 | Active Users (≥ 1x/month) |
| P4 | Man | 32 | Infrequent Users (A few times/year) |
| P5 | Man | 26 | New User (New < 6 months) |
| P6 | Woman | 29 | Active Users (≥ 1x/month) |

| | | | |
|---|---|---|---|
| P7 | Woman | 34 | Infrequent Users (A few times/year) |
| P8 | Man | 27 | Infrequent Users (A few times/year) |
| P9 | Woman | 31 | Active Users ($\geq$ 1x/month) |
| P10 | Woman | 28 | New User (New < 6 months) |

**Procedure**

*Usability* testing was conducted on-site *in* a controlled environment in October 2025. Each session lasted approximately 60 minutes and followed a standard protocol. Four task scenarios were selected to represent the main flow of Alodokter users: (1) Doctor Chat, (2) Appointment Booking, (3) Article Search, and (4) purchasing medication through Aloshop. These scenarios were chosen because they reflect the most frequent and crucial interactions on the telehealth platform. [14]. All tasks were conducted using the Alodokter app version 7.8.0 on an Android device to ensure consistency of the testing environment. Each participant completed all four tasks, yielding a total of 40 task-interaction data points for analysis.

Participants were asked to verbalize their thoughts during each task according to *the Think-Aloud Protocol* to capture cognitive processes, decision-making patterns, and emotional responses during the interaction. [15]. Before beginning, participants were given an explanation of the study's objectives and a brief session to familiarize themselves with the Think-Aloud technique, to ensure more consistent test implementation and reduce uncertainty during testing.

**Data collection technique**

Three types of data were collected: quantitative performance metrics, subjective usability ratings, and qualitative interaction data.

a. Performance Metrics. Quantitative performance data included task success and time-on-task. Each task was evaluated using a three-point success scale (0 = failed, 1 = succeeded with difficulty, 2 = succeeded without difficulty). Time-on-task measured the duration it took participants to complete the task from start to finish. Benchmark comparisons were made with similar health apps, namely Halodoc and Good Doctor, which were chosen because they have identical functions and are widely used in Indonesia [16].

b. Usability Score. Subjective satisfaction was measured using a four-point Likert scale (1 = very dissatisfied to 4 = very satisfied). The four-point scale was used to eliminate neutral responses and encourage more assertive assessments from participants. Observers simultaneously provided usability scores based on perceived ease of use, confidence levels, hesitation, and signs of frustration. The final satisfaction score was calculated as the average of the user and observer ratings, thus reflecting a balance between usability perceptions and observations [17].

c. Qualitative Interaction Data. Qualitative data were collected through direct observation, verbal Think-Aloud, and participant responses to probing questions. All sessions were screen- and audio-recorded with participant consent to capture behavior and verbal feedback comprehensively. A single-cycle open coding approach was used to generate initial codes, which were then grouped into broader themes representing usability issues, user frustrations, and positive interaction patterns [18].

d. The three-point task success scale (0 = failed, 1 = succeeded with difficulty, 2 = succeeded without difficulty) was adopted to capture varying levels of task completion quality rather than binary success outcomes. This approach has been widely used in task-based usability testing to reflect partial success and observable user difficulty during interaction, particularly in formative usability evaluations [19]

**e.** The four-point satisfaction scale was intentionally selected to eliminate neutral responses and encourage more decisive user judgments, a strategy commonly applied in usability and UX research to reduce central tendency bias. Similar forced-choice Likert scales have been employed in prior usability studies to obtain more apparent distinctions in perceived ease of use and satisfaction levels [17]

**Data analysis**

Data analysis was conducted based on the ISO 9241-11 usability framework, which evaluates effectiveness, efficiency, and satisfaction. Quantitative data obtained from task success scores, time-on-task measurements, and usability assessments were analyzed using descriptive statistics to identify performance patterns across the four user flows. Time-on-task values were reviewed for consistency, and extreme values were evaluated based on observed behavior rather than removed, to maintain the ecological validity of the dataset.

Qualitative data—including observation notes, Think-Aloud transcripts, and responses to probing questions—were analyzed using thematic analysis. [20]. Initial codes were developed from participants' behavioral and verbal accounts and then refined into high-level themes through iterative coding and affinity mapping. These qualitative themes were then triangulated with quantitative findings to strengthen interpretation and identify contextual factors underlying usability challenges. [21]

The integration of descriptive statistics and thematic analysis provides a comprehensive understanding of user performance, behavioral patterns, and interaction barriers across Alodokter's core service flows.

**Ethical Considerations**

All participants provided informed consent before participating in the study. They were informed of the study's purpose, confidentiality procedures, and the right to withdraw at any time. This study adhered to institutional ethical standards for human–computer interaction research. [22]

## Results And Discussion

In accordance with the research methodology outlined above, data collection was conducted offline with 10 (ten) participants (hereinafter referred to as P1 to P10) who met the criteria. The data from the Usability Testing session were systematically analyzed, combining quantitative and qualitative findings to address the research objectives.

**Quantitative Results**

Quantitative analysis evaluated three dimensions *of usability* —effectiveness, efficiency, and satisfaction— based on ISO 9241-11. Effectiveness was measured through *success rate,* efficiency through time-on-task, and satisfaction through usability scores from users and observers.

The quantitative assessment of effectiveness is summarized in Table 2. Among the four evaluations, Task 1 (Doctor Chat) demonstrated the best *usability,* with the highest success rate of 1.93. This indicates that users found the chat interface easy to understand and aligned with their expectations. Significantly different, Task 2 (Appointment Booking) recorded the lowest average score of 1.76. While Tasks 3 and 4 remained relatively stable with scores above 1.80, the significant drop in Task 2 indicates that the appointment booking process placed a higher cognitive load and was the biggest challenge in the application workflow.

**Table 2**. Average Task Success Score

| Task Flow | Average Success Score (0-2) |
|---|---|
| Task 1 – Doctor chat | 1.93 |
| Task 2 – Appointment Booking | 1.76 |
| Task 3 - Article | 1.81 |
| Task 4 - Aloshop | 1.83 |

To understand this complexity, Figure 1 details each participant's success rate on Task 2. The data confirms significant variation in user performance, explicitly indicating that participants P2 (1.33) and P7 (1.42) experienced the most important difficulty. These low scores indicate that some aspects within the appointment booking interface were potentially challenging for users.
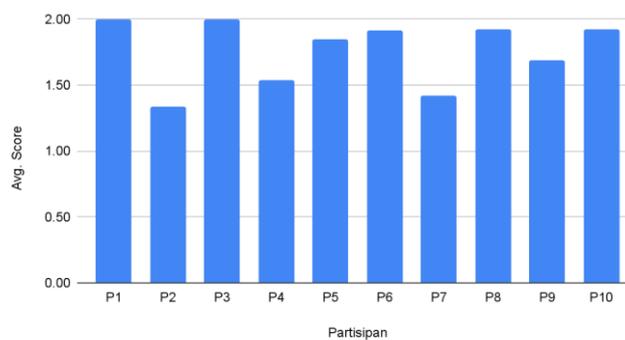


**Figure 1**. Average Success Score - Task 2 Appointment Booking

Efficiency performance was evaluated using *the time-on-task metric,* with comparisons to key competitor apps (Halodoc, Good Doctor, and MySiloam) as presented in Table 3. Overall, the results demonstrate a strong competitive advantage for Alodokter. Three of the four tested flows—Doctor Chat, Appointment Booking, and Article—were completed significantly faster than the competitor average. For example, the Appointment Booking task proved substantially more efficient, saving users significant time compared to industry standards. This finding suggests that the core navigation structure of these key features is well-optimized and supports rapid task completion.

However, the Aloshop flow stands out as a stark exception to this trend of efficiency. Unlike other tasks, the shopping scenario experienced significant delays, taking an average of 21 seconds longer to complete than the competitor's benchmark (1:39 vs. 1:18). This delay indicates a significant efficiency issue. It suggests that the current purchase path likely includes unnecessary steps or layout complexity that force users to spend more time and effort than they would on a more seamless experience on a competitor platform.

**Table 3**. Comparison of Average Task Completion Time

| Task Flow | Average Time (Alodokter) | Average Time (Competitor) | Results |
|---|---|---|---|
| Doctor Chat | 1:10 | 1:23 | Faster |
| Appointment booking | 4:16 | 5:46 | Faster |
| Article | 0:56 | 0:58 | Faster |
| Aloshop | 1:39 | 1:18 | **Longer** |

The final dimension, namely *Satisfaction,* was assessed through *subjective usability scores* on a scale of 1–4, as described in Table 4. An important finding from this data is the strong consistency between participant ratings and observer evaluations, which enhances the reliability of the feedback.

The Article and Doctor Chat features performed best in this category, receiving high user scores of 3.8 and 3.7, respectively. These scores indicate that users found the flow easy to use and helpful. Conversely, the Appointment Booking flow received the lowest subjective scores, with 2.5 from users and 2.3 from observers. This sharp decline supports previous findings regarding effectiveness; the various functional difficulties and obstacles users encountered during the booking process (as shown in Table 2) directly contributed to low satisfaction and negative perceptions of the feature's interface.

**Table 4**. Average Subjective Usability Score

| Task | Average User Score (scale 1-4) | Average Observer Score (scale 1-4) |
|---|---|---|
| Doctor Chat | 3.7 | 3.5 |
| Appointment booking | 2.5 | 2.3 |
| Article | 3.8 | 3.9 |
| Aloshop | 3.1 | 2.9 |

### Discussion of Qualitative Findings and Recommendations

Qualitative analysis based on observations, *the Think Aloud Protocol,* and *probing questions* identified several specific *pain points* that explain the previous quantitative findings. These qualitative findings provide deeper insight into the *usability challenges* of the Alodokter app. A summary of these findings, along with appropriate design recommendations, is presented in Table 5.

**Table 5**. Summary of Qualitative Findings and Recommendations for Improvement

| Task Flow | Location/Screen | Problem Findings and Feedback | Recommendations for Improvement |
|---|---|---|---|
| Doctor Chat | Doctor List Page | Users cannot sort doctors by lowest price. | Added sorting functionality by price and rating. |

| | | | |
|---|---|---|---|
| | Doctor Review Page | The review display is not intuitive, it is difficult to distinguish between positive and negative reviews. | Change the review system to a star scale (1-5) or differentiate the review icons. |
| | Chat Screen | The doctor's response felt too formal and stiff like AI. | Changed the auto-reply editor to be more humane and natural. |
| Appointment Booking | Overall Booking Flow | The medical procedure booking process was deemed too long and had too many steps (5 out of 10 Ps). | Simplify the booking flow by combining multiple steps into one screen. |
| | Doctor/Hospital Details Page | There is no crucial information regarding payment methods (especially insurance or BPJS cover). | Add insurance/BPJS cover information to the doctor and hospital details page. |
| | Doctor/Hospital Details Page | There is no estimated information or queue size for doctors/procedures on the selected day. | Displays the estimated number of queues on the selected day. |
| Article | Home Article Page | Users who are not sick expect trending/viral article content. | Added a "Trending Articles" section to the main page of the Articles feature. |
| Aloshop | Cart Page | Users (4 out of 10 Ps) expected to be able to return to the pharmacy storefront by clicking on the pharmacy name on the cart page, but did not find an entry point (link). | Added an entry point (link) to the pharmacy name on the cart page to return to the pharmacy display case. |
| | Cart & Checkout Flow | The process is inefficient; users have difficulty ordering other medications from the same pharmacy, leading to the risk of paying double shipping costs. | Provide a clear entry point in the cart to "Add to Order" from the same pharmacy before checkout. |

The proposed design recommendations can be further understood through established UX design principles. The simplification of the Appointment Booking process aligns with the principle of progressive disclosure, which suggests that information should be presented gradually to reduce cognitive overload. Displaying insurance and queue information directly on the doctor or hospital detail page supports visibility of system status, enabling users to understand constraints and expectations before proceeding. In the Aloshop flow, providing a clear entry point to add items from the same pharmacy reduces the risk of duplicate shipping costs and aligns with the principle of error prevention. Overall, these recommendations aim to reduce interaction cost by minimizing unnecessary navigation steps and improving decision efficiency across user flows.

To improve readability and provide a holistic overview of usability issues across different user flows, a cross-flow summary is presented. This summary synthesizes the main usability problems, affected usability dimensions, and key design recommendations identified from both quantitative and qualitative findings.

**Table 6**. Cross-Flow Summary of Usability Issues and Design Implications

| User Flow | Main Usability Issue | Affected Usability Dimension | Key Design Recommendation |
|---|---|---|---|
| **Doctor Chat** | Limited filtering and unclear reviews | Satisfaction | Add sorting and clearer review indicators |
| **Appointment Booking** | High cognitive load and missing contextual info | Effectiveness, Satisfaction | Simplify steps and display insurance & queue info |
| **Article Search** | Lack of engaging exploratory content | Satisfaction | Add "Trending Articles" section |
| **Aloshop** | Inefficient cart navigation | Efficiency | Enable adding items from same pharmacy |

This cross-flow summary highlights that while usability challenges vary across features, recurring patterns emerge, particularly related to cognitive load, information visibility, and interaction efficiency. The Appointment

Booking and Aloshop flows demonstrate the most critical usability concerns, suggesting that future design improvements should prioritize simplifying complex interactions and enhancing contextual information to support user decision-making.

**Table 7**. Comparison between SUS and Task-Based Usability Testing

| Aspect | System Usability Scale (SUS) | Task-Based Usability Testing (TBUT) |
|---|---|---|
| Data Type | Self-reported perception | Observed user behavior |
| Measurement Focus | Overall usability perception | Effectiveness, efficiency, satisfaction |
| Ability to Capture Errors | Limited | High |
| Insight into Cognitive Load | Indirect | Direct (via Think Aloud) |
| Suitability for Complex Flows | Moderate | High |
| Contextual Interaction Analysis | Low | High |

While the System Usability Scale (SUS) is effective for capturing users' overall perceptions, it provides limited insight into actual interaction behavior and task-level difficulties. In contrast, Task-Based Usability Testing (TBUT) enables direct observation of user performance, decision-making processes, and cognitive challenges during real task execution. This makes TBUT particularly suitable for evaluating complex, multi-step interactions and cross-flow transitions within hybrid mHealth applications, where usability issues often emerge during task execution rather than post-use reflection. Based on quantitative and qualitative data, a summary of the main *usability issues* for each user flow is as follows:

In the **"Chat with Doctor" flow,** quantitative results indicated high Effectiveness and Satisfaction scores, but qualitative findings revealed minor Satisfaction issues. Users mentioned being unable to sort doctors by price and having difficulty understanding the doctor review system. Additionally, some participants felt that the chat responses from doctors felt "too formal" and "AI-like." Key recommendations included adding sorting and rating features, improving the review system's interface, and restructuring the response style to sound more human and empathetic.

In the Appointment Booking flow, quantitative data consistently showed it was the most problematic, with the lowest *Success Rate* (1.76) and the lowest Satisfaction score (2.5). Qualitative findings explained why: the booking process was perceived as too long and repetitive, consisting of too many steps (reported by 5 out of 10 participants). Other critical *pain points* were the lack of essential information—particularly payment methods and insurance/BPJS coverage—and the lack of queue estimates for selected doctors or services. Therefore, a key recommendation is to simplify the booking process by consolidating steps into a single screen and displaying payment and queue information directly on the doctor/hospital detail page.

From a cognitive perspective, the low usability performance of the Appointment Booking flow can be explained through the lens of cognitive load theory. The multi-step booking process requires users to process excessive information across several screens, including doctor selection, schedule confirmation, payment options, and service details. This fragmented structure increases extraneous cognitive load, forcing users to retain information from previous steps while making decisions in subsequent screens. Additionally, the absence of critical contextual information, such as insurance or BPJS coverage and queue estimation, creates decision complexity, as users must make choices without sufficient decision support. These conditions contribute to higher mental effort, hesitation, and task difficulty, which were reflected in the lower success and satisfaction scores for this flow.

In the Article flow, qualitative findings revealed unmet user needs. While this flow achieved a high satisfaction score (average 3.8), some participants stated that they wanted to read trending health articles even if they weren't searching for specific medical information. Therefore, a key recommendation is to add a "Trending Articles" section to the main Article page to increase engagement and content accessibility.

Finally, in the Aloshop flow, which quantitatively demonstrated issues in the Efficiency dimension, it was found to be 21 seconds slower than the competition. Qualitative findings explain why: four participants expected to be able to add another item from the same pharmacy directly from the cart page, but found no access point to do so. This forced users through an inefficient flow and potentially resulted in double shipping costs if they accidentally purchased from a different pharmacy. The recommendation is to add an access point (link) to the pharmacy name on the cart page to return users to that pharmacy easily.

These findings are consistent with previous research by [23] Which also highlighted *usability* and navigation issues in telemedicine apps in Indonesia. However, this study makes a novel contribution by identifying specific *pain points* and data-driven design recommendations across four core flows of the Alodokter app—specifically, issues related to insurance/BPJS visibility in Appointment Booking and inefficiencies in the Aloshop checkout process. These aspects have not been previously documented in digital health research publications in Indonesia.

## Conclusion

This study evaluated *the usability* of the Alodokter application using *Task-Based Usability Testing,* as defined in ISO 9241-11. The results showed that the Doctor Chat and Article Search flows had the best *usability performance* across effectiveness, efficiency, and satisfaction. Conversely, the Appointment Booking flow was the most problematic due to a long, repetitive interaction process and a lack of important information such as payment methods and BPJS/insurance coverage. The Aloshop flow also showed efficiency issues, as evidenced by a longer checkout duration than that of *benchmark* competitor applications. These findings indicate that although the core service features already provide a good experience, Alodokter still needs to simplify the flow and improve visuals to ensure consistent usability across all key features. From a theoretical and methodological perspective, this study contributes to the usability research literature by demonstrating the applicability of Task-Based Usability Testing (TBUT) in evaluating hybrid medical–commercial mHealth platforms within the Indonesian context. Unlike perception-based approaches, TBUT enabled the identification of cross-flow usability issues that emerge when users transition between healthcare services and commercial features within a single application ecosystem. This study extends prior mHealth usability research by empirically linking task performance metrics with cognitive and interactional challenges across distinct service flows, thereby offering a more nuanced understanding of usability in integrated digital health platforms.

This study has several methodological limitations that should be considered when interpreting the findings. First, the sample size consisted of only ten participants. Although this number aligns with established usability testing guidelines, it limits the statistical generalizability of the results beyond formative usability insights. Second, all participants were recruited from the Greater Jakarta (Jabodetabek) area, which may not fully represent users from other regions in Indonesia who may differ in terms of healthcare access, digital literacy, and usage contexts. Third, usability testing was conducted exclusively on the Android platform; therefore, interaction patterns, interface behaviors, and potential usability issues specific to other platforms, such as iOS, were not captured. As a result, the findings should be interpreted as exploratory rather than definitive and may not be directly generalized to broader user populations or alternative platform environments. This study has several methodological limitations that should be considered when interpreting the findings. First, the sample size was only 10 participants, which, although consistent with usability testing guidelines, limits the statistical generalizability of the results. Second, all participants were recruited from the Greater Jakarta (Jabodetabek) area, which may not fully represent users from other regions with different healthcare access patterns or digital literacy levels. Third, usability testing was conducted exclusively on the Android platform; therefore, interaction patterns and usability issues on other platforms, such as iOS, were not captured. These limitations suggest that the findings should be interpreted as exploratory rather than definitive and highlight the need for future studies involving larger, more geographically diverse samples and multi-platform testing environments. Practically, the findings provide actionable design insights for UX designers, product managers, and developers of digital health applications. The identified usability issues particularly those related to cognitive load in Appointment Booking and interaction inefficiencies in Aloshop highlight the importance of simplifying multi-step processes, improving visibility of contextual information, and reducing unnecessary interaction costs. The design recommendations proposed in this study can serve as practical guidelines for improving user experience, trust, and engagement in hybrid healthcare platforms, especially in emerging digital health markets such as Indonesia.

## References

[1]     R. A. Silvalena, S., & Syakurah, "Utilization and perception of telemedicine in Indonesia: Healthcare providers' point of view," *Jurnal Kesehatan Masyarakat Andalas*, vol. 16, no. 1, 2022.

[2]     A. M. Deniz-Garcia, A., Fabelo, H., Rodríguez-Almeida, A. J., Zamora-Zamorano, G., Castro-Fernández, M., Del Pino Alberiche Ruano, M., Solvoll, T., Granja, C., Schopf, T. R., Callico, G. M., Soguero-Ruiz, C., & Wägner, "Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: Current scenario and challenges," *Journal of Medical Internet Research*, vol. 25, 2023.

[3]     H. Alfian, S. D., Fudholi, A., Nur, S., & Sukandar, "Evaluation of usability and user feedback to guide telepharmacy application development in Indonesia: A mixed-methods study.," *Frontiers in Digital Health,* vol. 6, no. 1, 2024.

[4]     X. Wang, Q., Zhang, R., Xue, B., Hu, J., Wang, W., Zhang, Y., & Li, "Usability evaluation of mHealth apps for elderly individuals," *BMC Medical Informatics and Decision Making*, vol. 22, no. 20, 2022.

[5]     R. Galavi, Z., Montazeri, M., & Khajouei, "Which criteria are important in usability evaluation of mHealth applications: An umbrella review," *BMC Medical Informatics and Decision Making*, vol. 24, no. 365, 2024.

[6]     H. Philip, S., & Hidayaturrahman, "Effective approaches for user engagement improvement in mobile health applications: A comprehensive literature analysis. Engineering," *Engineering, Mathematics and Computer Science Journal (EMACS)*, vol. 6, no. 3, pp. 185–191, 2024.

[7]     H. Dyayu, A. L., Beny, B., & Yani, "Evaluasi usability aplikasi PeduliLindungi menggunakan metode usability testing dan SUS," *Jurnal Manajemen Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 395–404, 2023.

[8]     P. Weichbroth, "Usability testing of mobile applications: A methodological review.," *Applied Sciences*, vol. 14, no. 5, 2023.

[9]     C. Rochat, J., Ehrler, F., Siebert, J., Ricci, A., Garretas Ruiz, V., & Lovis, "Usability testing of a patient-centered mobile health app for supporting and guiding the pediatric emergency department patient journey: Mixed methods study," *JMIR Pediatrics and Parenting*, vol. 5, no. 1, 2022.

[10]    A. Tsimicalis, A., Davidson, L., Saxena, S., Stasiulis, E., Riazi, A., Laflamme, A., Hervieux, V., & Ferland, "The design, development, and usability testing of an eHealth program for youths with osteogenesis imperfecta: Protocol for a 2-phase user-centered mixed methods study," *JMIR Research Protocols*, vol. 12, 2023.

[11]    C. Rochat, J., Ehrler, F., Siebert, J., Ricci, A., Garretas Ruiz, V., & Lovis, "Usability testing of a patient-centered mobile health app for supporting and guiding the pediatric emergency department patient journey: Mixed methods study," *JMIR Pediatrics and Parenting*, vol. 5, no. 1, 2022.

[12]    J. Nielsen, "How many test users in a usability study? Nielsen Norman Group".

[13]    A. Tsimicalis, A., Davidson, L., Saxena, S., Stasiulis, E., Riazi, A., Laflamme, A., Hervieux, V., & Ferland, "The design, development, and usability testing of an eHealth program for youths with osteogenesis imperfecta: Protocol for a 2-phase user-centered mixed methods study," *JMIR Research Protocols*, vol. 12, 2023.

[14]    M. Septiani, W., Rahmawati, N., Safitri, D. M., & Luis, "Usability evaluation for mobile health application: Systematic literature review," *SINERGI (Indonesia)*, vol. 28, no. 2, pp. 287–304, 2024.

[15]    A. Wardani, P. T., & Pratama, "Analisis usability menggunakan metode Think Aloud dan Heuristic Evaluation pada aplikasi JMO (Jamsostek Mobile).," *SISFO: Jurnal Ilmiah Sistem Informasi*, vol. 7, no. 1, pp. 19–36, 2023.

[16]    S. P. Afrizal, S., Prihatni, R., & Hastono, "Evaluasi kualitas aplikasi kesehatan menggunakan System Usability Scale," *Jurnal Manajemen Informasi Kesehatan*, vol. 9, no. 1, pp. 1–9, 2024.

[17]    H. Dyayu, A. L., Beny, B., & Yani, "Evaluasi usability aplikasi PeduliLindungi menggunakan metode usability testing dan SUS," *Jurnal Manajemen Teknologi dan Sistem Informasi*, vol. 3, no. 1, pp. 395–404, 2023.

[18]    R. A. Silvalena, S., & Syakurah, "Utilization and perception of telemedicine in Indonesia: Healthcare providers' point of view," *Jurnal Kesehatan Masyarakat Andalas*, vol. 16, no. 1, 2022.

[19]    F. Khajouei, R., & Farahani, "A combination of two methods for evaluating the usability of a hospital information system.," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 84, 2020.

[20]    V. Braun, V., & Clarke, "Thematic analysis: A practical guide. SAGE Publications," 2021.

[21]    C. Crehan, "Optimising and evaluating usability, acceptability and feasibility of a neonatal digital health system in a low-resource setting: The NeoTree case study (Doctoral dissertation, University College London).," 2023.

[22]    V. Braun, V., & Clarke, "Thematic analysis: A practical guide. SAGE Publications," 2021.

[23]    R. Setianingsih, S., Muyarman, B., Akmalia, S. M., Lim, J. S., Nolasco, J. J., & Widayanti, "Evaluasi tingkat usability aplikasi Halodoc menggunakan pengukuran System Usability Scale.," *Journal of Multidisciplinary Science*, vol. 1, no. 12, pp. 778–785, 2024.