ISSN 2355-0082

Implementation of Ergonomic Criteria for Evaluating the Logistic Mobile App User Interface Design

Hany Alexandra [1], Muhammad Ulil Amri [2], Fadhilah Eka Putri[3], Andi Jamiati Paramita [4],
Rena Nainggolan[5], Fenina Adline Twince Tobing[6]
[1,2,3,4]Sistem Informasi, Institut Teknologi dan Bisnis Kalla, Makassar, Indonesia
[5]Universitas Methodist Indonesia, Medan, Indonesia
[6]Universitas Multimedia Nusantara, Tangerang, Indonesia
[1]hanyalexandra@kallainstitute.ac.id

Accepted 12 January 2026
Approved 11 February 2026

Abstract*— The rapid growth of digital logistics has increased dependence on mobile delivery applications to manage shipments, package information, and courier operations under strict time constraints. In such environments, ergonomic and usable user interface (UI) design is essential to reduce cognitive load, minimize errors, and maintain operational efficiency. However, empirical usability research focusing specifically on delivery service applications remains limited, particularly studies conducted in realistic usage contexts. This study evaluates the ergonomic quality of a mobile delivery application using online usability testing. A quantitative evaluation framework is applied, employing three established metrics: the System Usability Scale (SUS), task completion time, and error rate. The results indicate an average SUS score of 57.7 out of 100, placing the application in the *marginal usability* category. Only 25.0% of participants rated the application as acceptable (SUS ≥ 68), while 43.8% rated it as marginal and 31.3% as poor. Task-based testing showed that although 100% of participants completed the primary navigation task, 56.3% reported experiencing difficulty, indicating issues related to navigation clarity and interaction flow. In contrast, 25.0% of users failed to correctly complete the invoice-related task, reflecting weaknesses in information hierarchy and labeling. Based on these findings, the study proposes data-driven UI design improvements aimed at enhancing navigation structure, information clarity, and cognitive ergonomics. The novelty of this research lies in integrating online usability testing with ergonomic evaluation in the delivery application domain, providing scalable quantitative benchmarks and practical design guidance for improving the usability of mobile logistics applications.

***Index Terms*— Application; Design; Ergonomic; Mobile; UI**

I. Introduction

The growth of digital logistics services has prompted the adoption of mobile delivery apps to track shipments, handle package information, and support courier operations. These apps are critical to maintaining the speed, precision, and dependability of the delivery process, especially as e-commerce volumes grow globally. In such a high-demand setting, effective user interface (UI) design is critical for lowering cognitive load, increasing efficiency, and reducing operational error [1]. Recent research suggests that design methods should focus on improving the user interface and interactions of mobile applications to make them more suited to the community [2], which suggests that better interface and interaction might improve the purpose of the application. Therefore, design should begin with a clear understanding of the community's characteristics and usage context (e.g., digital literacy, language preferences, accessibility needs, device limitations, and time pressures), as these factors shape how users interpret interface elements and complete tasks or scenarios presented in the testing phase [3]. In practice, this typically translates into a user-centered, iterative workflow, such as contextual inquiry, task analysis, prototyping at low to high levels of detail, and iterative usability testing to identify friction points in navigation, information clarity, and user feedback [4]. When these interaction friction points are reduced, users tend to make fewer errors, complete tasks more reliably, and require less effort to achieve their goals, which can increase adoption and continued use [5]. However, improvements should be validated with measurable outcomes (e.g., task completion rates, task time, error rates, and perceived usability) to ensure that interface changes truly support the app's intended purpose, rather than simply enhancing visual appeal [6].

Although research on mobile usability and digital ergonomics has expanded significantly over the last decade, most studies concentrate on broader domains such as e-commerce platforms [7], mobile banking, or navigation-based applications [8]. Furthermore, empirical research on delivery service applications is scarce. Delivery apps have distinct and time-sensitive processes that need users, including customers and delivery service providers, to accomplish tasks promptly and properly [9]. Because of the domain-specific complexity, traditional usability findings from other mobile applications may not be completely applicable to delivery-oriented systems, revealing a substantial research need.

Furthermore, much research is irrelevant to actual situations. These findings do not accurately reflect real-world contextual elements that influence user behavior. Recent research emphasizes the growing necessity for remote and online usability assessment methodologies that collect data in natural usage scenarios [10]. Online testing enables researchers to realistically watch user interactions and measure performance measures such as task completion time, error rates, and usability judgments without restriction [11]. However, the use of online usability testing, particularly for investigating ergonomic requirements in delivery apps, is still underexplored in the 2020-2025 research landscape.

These gaps raise several key questions:

(1) How well does the UI of a delivery application align with modern ergonomic and mobile usability principles?

(2) What usability issues? Such as prolonged task time or increased error frequency—occur during user interaction?

(3) Can quantitative metrics such as the System Usability Scale (SUS), task-time efficiency, and error rate reliably represent the ergonomic quality of the delivery interface?

(4) What UI improvements can be recommended to enhance ergonomic suitability and user performance?

Based on these questions, this study aims to evaluate the ergonomic quality of a mobile delivery application using online usability testing. Three quantitative usability metrics are used: SUS score [12], task completion time, and error rate; each of which provides valuable insights into interface performance and ergonomic suitability [13]. The study also identifies common challenges faced by users during delivery-related tasks and proposes design improvements to enhance interface clarity, efficiency, and cognitive ergonomics [14].

The novelty of this research is expected to be useful due to its ongoing and renewable contribution, building upon previous research. Online usability testing and user ergonomic heuristics are included in some of this research, a methodological combination that is uncommon in delivery-oriented UX research between 2020 and 2025. Many other studies provide performance-based evaluation models tailored to delivery processes to solve problems that remain underrepresented in current human-computer interaction (HCI) research [15]. Usability measurements are scalable and repeatable, which will make comparing and benchmarking future logistics mobile apps easier [16]. By providing practical UI design suggestions based on actual data, the result seeks to assist clients and logistics operators in developing more efficient, user-friendly, and ergonomic delivery service applications [17].

The research questions in this study were formulated based on three interrelated considerations. First, prior studies on mobile usability and digital ergonomics have predominantly focused on general application domains, leaving delivery service applications underexplored, particularly in terms of quantitative ergonomic evaluation. Second, the study adopts a performance-oriented usability perspective, requiring research questions that can be examined using measurable indicators rather than purely qualitative judgments. Third, the research questions were aligned with the selected evaluation metrics—System Usability Scale (SUS), task completion time, and error rate—to ensure a direct link between the identified research gap, data collection methods, and analytical outcomes. Consequently, the research questions were designed to (1) assess the degree of ergonomic alignment of the delivery application interface, (2) identify usability problems manifested through efficiency and error patterns during task execution, and (3) determine how effectively quantitative usability metrics represent the ergonomic quality of the interface and inform design improvement recommendations.

II. Method

a) *Research Design*

This study used a quantitative descriptive-evaluative design with a task-based usability testing approach and expert-based ergonomic evaluation. Ergonomics evaluation assesses the extent to which a predefined logistics mobile application's user interface (UI) meets ergonomic criteria, whereas online usability testing measures actual user performance and perceived usability [18], using quantitative metrics such as System Usability Scale (SUS) scores, task completion times, and error rates [19].

The overall goals of the predefined methods are to evaluate UI approaches in accordance with established ergonomic and mobile usability principles [20], identify usability issues that arise during interactions with delivery-related tasks [21], and make empirically supported interface improvement recommendations.

The research involved focuses on consumer-facing mobile logistics apps that provide essential delivery services such as shipment monitoring, package information management, order creation, and customer care. Customers often use these applications, which are installed on Android and iOS smartphones, to track and manage package deliveries in real-time.

This study focuses on the customer-side interface, which is directly involved in tasks like shipment tracking, changing delivery information, and submitting delivery requests. This research also focuses on other stakeholder interfaces (for example, couriers and administrators).

b) *Participants Gathering*

The study's demographics featured active users across different logistics mobile apps. The criteria specified in the study are as follows:

1. Aged 18 to 30 years old.
2. Willing to use the specified apps at least three times in the last three months for delivery-related activities (for example, shipment tracking, package delivery).
3. To access the apps, use your smartphone (Android or iOS) as the primary device.
4. Capable of following written instructions in the language used for the study.

Participants were recruited using a non-probability purposive sample strategy [22], mostly via online channels such as messaging apps, social media, and app-specific user groups.

The participant age range was limited to 18–30 years to ensure sample homogeneity and to minimize variability related to age-dependent differences in mobile technology usage. Individuals within this range represent the most active segment of mobile application users and generally demonstrate higher levels of smartphone familiarity, faster adaptation to interface changes, and more consistent interaction patterns with app-based services. Restricting the sample to this age group reduces potential confounding effects associated with age-related cognitive decline, perceptual differences, or technology anxiety that may influence task performance and usability perceptions. Consequently, this boundary improves the internal validity of the usability evaluation, while the findings are intentionally interpreted as representative of active mobile logistics application users within this age group rather than the general population.

For online usability testing, a minimum of 20–30 participants is recommended to identify common usability issues and obtain stable descriptive metrics. To increase the reliability of SUS scores and performance indicators, the study targeted approximately 20-30 participants according to related research whose data met the inclusion criteria and passed basic data quality checks (e.g., completion of all tasks and questionnaires and adherence to established criteria).

c) Variables and operational definitions

The primary constructs explored in this study are:

1. Implementation of Ergonomic Criteria (Independent Construction)

a) Refers to how well the specified logistics application's user interface corresponds to ergonomic and usability criteria. This design was operationalized using an expert-based ergonomics checklist that included various dimensions, such as:

b) Visual ergonomics (legibility, contrast, font size, and color utilization).

c) Layout and navigation: consistency, grouping, and information hierarchy

d) Interaction controls (touch target size, spacing, and reach).

e) Feedback and error management (system response, error messages, and confirmations)

f) Cognitive workload (number of steps, information density, label clarity)

3. Each checklist item is assessed using a Likert scale (1 = not fulfilled, 5 = fully satisfied). A higher overall score suggests a more advanced level of ergonomics implementation.
4. The 10-item System Usability Scale was used to evaluate tester perceptions of usability. Each question was also scored on a 5-point Likert scale, from "strongly disagree" to "strongly agree." The scores were translated to the normal SUS scale (0–100). Higher SUS ratings represent improved perceived usability.
5. Task completion time is the amount of time (in seconds) taken by a participant to complete a given task scenario, calculated from the moment the task begins until it is successful or failed. Time is collected for each task and averaged across all participants.
6. The error rate is the frequency with which participants make noticeable errors when performing a task, such as incorrectly navigating to an unexpected section of the screen, failing form submissions, or completing activities incompletely or inaccurately. The error rate can be represented as follows:
   a) Mean number of errors per participant per task, or
   b) Percentage of participants that fail to complete a task properly.

d) Instruments

Several instruments are employed in this study:

1. **Ergonomic Criteria Checklist**

The arranged checklist is based on ergonomic and application usability concepts gleaned from applicable standards and prior research. It addresses several factors (visual ergonomics, layout and navigation, interaction controls, feedback and error handling, and cognitive load). Each item is scored on a scale of 1 to 5. At least one ergonomics or user experience professional reviews the checklist to confirm that the material is valid.

2. **Task scenarios for usability testing**

A set of task scenarios is prepared to reflect core delivery-related activities. Examples of tasks include:

a) Task 1: Track the current status of a specific shipment based on a given tracking number.

b) Task 2: Update the delivery address for an ongoing shipment.

c) Task 3: Create a new delivery order from a specified pick-up address to a destination address.

d) Task 4: Access and use the customer support/contact feature regarding a delayed package.

Tasks are written in clear, concise instructions and are presented to participants in the same order.

3. **Usability Scale (SUS) questionnaire**

The SUS questionnaire consists of 10 standard items measuring perceived usability of the application as a whole. After completing all tasks, participants are asked to fill in the SUS questionnaire based on their overall experience during the session.

4. **Demographic and background questionnaire**

A brief questionnaire was utilized to collect demographic information (age, gender) as well as use information (app experience, frequency of use, user knowledge with mobile delivery applications in general).

5. **Task performance recording sheet**

A recording sheet or digital form is used to capture task completion time and error occurrences for each participant and each task. When possible, screen recording tools or analytic logs may be used to increase accuracy in measuring time and errors.

Data collection is carried out in several stages:

1. **Preparation phase**

a. Conduct a literature review on ergonomic criteria, mobile usability, and logistics-related mobile applications.

b. Create an ergonomic criterion checklist and task scenarios.

c. Create web forms for consent, task instructions, SUS, and demographic information.

d. Conduct a pilot test with 3-5 participants to ensure that task instructions are clear, the data collection platform is user-friendly, and the overall duration is reasonable.

2. **Ergonomic evaluation**

a. Selected logistics applications were systematically evaluated using an ergonomic checklist.

b. Researchers examined interface screens and interaction flows that corresponded to predetermined tasks.

c. Each checklist item was assessed, and the scores were aggregated to obtain a total ergonomic score for each dimension.

3. **Online usability testing**

a. Participants are instructed to perform the tasks on their own smartphones in their natural environment.

b. Task completion time and errors are recorded, either through:

i. self-report with clear timing guidelines, or

ii. remote observation or screen recording (if the platform allows).

c. After completing all tasks, participants fill in the SUS questionnaire and any additional questions regarding their experience.

1. **Data cleaning and validation**

a. Responses with incomplete tasks, missing SUS data, or inconsistent answers are identified.

b. Only data that meet the inclusion criteria and completeness requirements are retained for analysis.

Data analysis consists of descriptive and analytic steps:

1. **Descriptive analysis**

a. Calculate the mean and standard deviation of scores for each ergonomic dimension, as well as the total ergonomic score for the application.

b. Compute the SUS scores obtained based on the results of evaluation tests per individual and obtain the mean, standard deviation, and distribution of SUS scores. Interpret scores based on commonly used SUS acceptance ranges (Kaya et al., 2019).

c. Determine the mean, median, and standard deviation of completion times for each task.

d. Calculate the mean number of errors per task and the proportion of participants who failed or experienced difficulty in each task.

2. **Analytic interpretation of ergonomic implementation and performance metrics**

a. Compare ergonomic checklist scores based on heuristic evaluation & usability-based scenario (Andika, 2024)

b. Identify patterns linking specific ergonomic issues (e.g., small touch targets, low contrast, complex navigation, element misconceptions) with observed usability problems.

e) Research Phases

Figure 1 illustrates the overall research framework adopted in this study, which structures the research process into nine sequential phases grouped into five major clusters. The figure provides a visual overview of how the study progresses from initial problem conceptualization and research scoping, (1) Phases 1-2 focused on issue conceptualization, research scope determination, and research objective establishment. (2) Phase 3 was responsible for sampling design, including determining the target population and sampling frame, inclusion-exclusion criteria, sampling technique, and respondent recruitment approach. (3) Phases 4-5 formed the requirements definition group, which transformed insights into organized task cases and scenarios with explicit objectives and success criteria for each case. Once the specifications were completed, (4) Phases 5-7 included data gathering and knowledge analysis, in which information was methodically gathered and analyzed to find patterns, gaps, and consequences that corresponded to the preset scenarios. The last cluster, (5) Phases 8-9, integrated the general interpretation of the preceding phases' results and generated final recommendations, including priorities and practical actions based on the examined data.

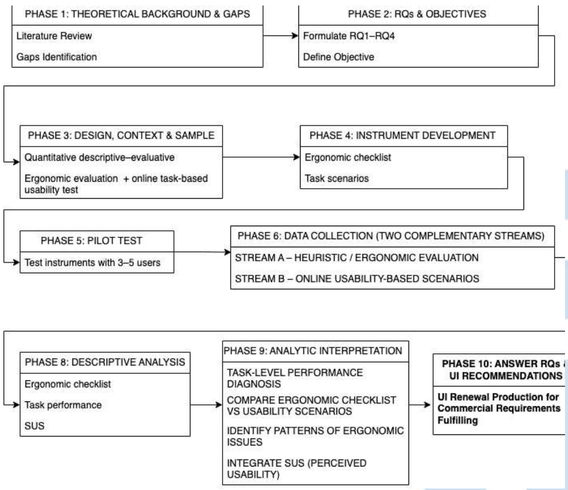

Fig 1. Research Phase

III. Result and Discussion

The research study's data were subsequently processed using batch-based interpretation in nine preset processes. This enabled the first contextual findings to be gradually transformed into practical needs, empirical evidence, and suggestions. All findings are split into several batches. Each batch is related to the other batch within the exact phase to complete the contextual purpose of the research.

A. Batch 1: Phases 1–2 (initial exploration and outlining)

This section presents preliminary contextual data and a project framework. In this section, Ulil Trans is compared to comparable apps using App Store/Google Play ratings as a high-level sentiment indicator, indicating how the app could be viewed in the larger market. However, because shop evaluations are impacted by numerous aspects other than interface design (such as stability, performance, and service quality), this comparison is viewed as supporting context rather than direct evidence of UI/UX excellence.

B. Batch 2: Phase 3 (design sampling and participant recruitment)

Phase 3 determined the sample design and defined the demographics of the usability testers, including the participant age restriction (18-30 years) as the primary demographic control, and limited participation to active Ulil Trans mobile app users. Age was purposefully chosen as the only demographic variable to preserve sample comparability and limit variability caused by age-related variations in smartphone familiarity and engagement patterns with mobile apps. In this study, active users were identified as those who had used the Ulil Trans app at least once during the previous 30 days, ensuring that task performance and SUS replies were based on recent, real use rather than first-time exploration. As a result, future interpretation of the data is confined to the experiences of active users within the designated age range, whereas onboarding and churn barriers among first-time users or users who quit usage of the app are beyond the scope of this phase.

Demographic points:

1. Primary demographic variable: Age only
2. Age range boundary: 18–30 years
3. Additional demographic segmentation: Not applied (e.g., gender, education, occupation were not used) to avoid sample fragmentation
4. User status boundary: Active users only
5. Active user definition: Used the Ulil Trans app at least once within the last 30 days
6. Interpretation boundary: Findings represent active Ulil Trans users aged 18–30; not generalizable to users outside the range or to first-time/lapsed users
7. Recruitment basis: Participants selected based on the anticipated user profile (rather than random selection), assuring relevance for task-based usability testing.

Phase 3 additionally included a brief UI picture of Ulil Trans to help with demographic borders and offer context for upcoming job situations. This temporary UI snapshot served as the study's baseline, exhibiting the interface state used and introduced to participants throughout testing and assessment, as well as defining the primary screens and navigation features essential to the task scenarios tested in the following phases.

C. Batch 3: Phase 4 to 5 (determination of requirements based on task cases and scenarios)

In this phase, all requirements have been integrated into selected questions with suitable requirements that have been concluded based on the previous phase or stages.

1. **Demographic Data & Experience**
   a) Age
   b) Frequency of Ulil Trans Use
      (a) 1–2 times
      (b) 3–5 times

(c) 6–10 times

(d) More than 10 times

(e) Never used

c) Frequency of Other Logistics Apps

(a) Rarely / almost never

(b) Sometimes (1–2 times per month)

(c) Fairly often (3–5 times per month)

(d) Very often (>5 times per month)

2. **Task 1: Navigating the Menu Options on the Home Page**

a) Did you complete all the steps in Task 1 (opening the Ticket Booking menu, returning to the homepage, then opening the Notifications menu)?

(a) Yes, I managed to complete all the steps without any difficulty.

(b) Yes, I managed to complete them, but with some difficulty.

(c) No, I did not manage to complete all the steps.

b) Estimated time required to complete this task

(a) < 1 minute

(b) 1 minute

(c) 1 minute 30 seconds

(d) 2 minutes

(e) > 2 minutes

c) While performing Task 1 (opening the Ticket Booking menu, returning to the homepage, then opening the Notifications menu), what difficulties did you encounter? (You may select more than one)

(a) I'm having trouble finding the Ticket Booking menu on the homepage.

(b) I'm confused by the menu icons or labels (I don't immediately understand what the icons/text mean).

(c) I'm having trouble finding how to return to the homepage after opening a menu.

(d) I'm having trouble finding the Notifications menu or I'm not sure where the information is located.

(e) I find the navigation between menus unclear, making it take a long time to navigate from one section to another.

(f) Other

3. **Task 2: Order History**

a) Were you successful in finding and understanding the order status you were looking for on the History page?

(a) Yes, I managed to complete all the steps without any difficulty.

(b) Yes, I managed to complete them, but with some difficulty.

(c) No, I was unable to find or understand the order status I was looking for.

(d) Other

b) Estimated time required to complete this task

(a) < 1 minute

(b) 1 minute

(c) 1 minute 30 seconds

(d) 2 minutes

(e) > 2 minutes

c) While viewing the History page and trying to understand the information on the order list, what difficulties did you encounter? (You may select more than one)

(a) I had difficulty finding a specific order.

(b) The information in each history (name, origin-destination, service type, status) felt too dense and difficult to read quickly.

(c) I didn't understand the meaning of the colors used to indicate trip status.

(d) I had difficulty distinguishing the type of service (freight vs. passenger ticket) from the display.

(e) I had difficulty scrolling or navigating through the History page to view older or newer orders.

(f) No difficulty.

(g) Other

d) What was the most confusing or hindering aspect of this task?

4. **Task 3: Booking Passenger Tickets**

a) Were you able to easily fill out the passenger ticket booking form?

(a) Yes, it worked.

(b) No, it didn't work.

b) Estimated time required to complete this task

(a) < 1 minute

(b) 1 minute

(c) 1 minute 30 seconds

(d) 2 minutes

(e) > 2 minutes

c) What problems did you encounter while performing this task? (You can select more than one)

(a) I was able to fill in all the data easily without any issues.

(b) I was confused about the order or location of the required fields.

(c) The text/labels in the fields were unclear or too small, making them difficult to read.

(d) I entered data incorrectly and didn't receive a clear error message.

(e) Buttons or elements that needed to be touched (e.g., the continue/submit button) were difficult to find or difficult to touch.

(f) Other

d) What was the most confusing or hindering aspect of this task? (You can leave blank if none occurred)

5. **Task 4: Checking Invoices**

a) Do you understand the details of the invoice listed on the page?

(a) Yes, it worked.

(b) No, it didn't work.

b) Estimated time required to complete this task

(a) < 1 minute

(b) 1 minute

(c) 1 minute 30 seconds

(d) 2 minutes

(e) > 2 minutes

c) While viewing the invoice page, what were your experiences? (You can choose more than one)

(a) I had difficulty understanding some of the information on the invoice (e.g., terms, labels, or transaction descriptions).

(b) The information felt too dense, making it take me a long time to find the details I was looking for.

(c) The text or number display (date, amount, transaction code) was too small/difficult to read.

(d) I had difficulty finding the buttons or menus to download the invoice (download/print/save).

(e) I was confused by the layout of the information (the order details, total payment, and shipping status were unclear).

(f) Other

d) What was the most confusing or hindering thing about completing this task?

6. **Usability Questionnaire (SUS)**

a) I think I'll use Ulil Trans frequently.

b) I find Ulil Trans too complicated to use.

c) I find Ulil Trans easy to use.

d) I feel like I need help from others to use Ulil Trans.

e) I find Ulil Trans' features well-organized.

f) I feel like there are too many inconsistencies in Ulil Trans.

g) I imagine most people will quickly become proficient with Ulil Trans.

h) I find Ulil Trans complicated and confusing.

i) I feel confident using Ulil Trans.

j) I need to learn a lot before I can use Ulil Trans well.

7. **Additional Open-Ended Questions**

What part of the Ulil Trans interface or flow do you find most helpful and user- friendly? Please describe it briefly.

D. Batch 4: Phase 5 to 7 Results (Pilot Test, Data Collection, Descriptive Analysis)

Based on the data provided, the usability testing of the Ulil Trans application generated moderate but inconsistent results, with an average SUS score of 57.7/100 (marginal category). Only 25.0% of respondents rated the usability as "good/acceptable" (SUS ≥ 68), while 43.8% rated it as "marginal" (SUS 50-<68) and 31.3% rated it as "poor" (SUS < 50). At the task level (n = 32), users were generally able to complete the task. In Task 1 (homepage navigation: Ticket Booking → Home → Notifications), 100% of the testers were successful, but the majority did not succeed (56.3%) stating that they succeeded with some difficulties, which indicates problems with the clarity of the navigation structures, navigation flows, and visual clues while navigating menus; in Task 2 (finding and understanding booking status in History), most of the testers completed the scenario without difficulty. In Task 4 (understanding invoice details and downloading invoices), 75.0% stated that the majority of testers understood the flow and procedures for downloading invoices, while 25.0% did not. This was indicated by indicators that showed weaknesses in information design, such as display hierarchy, cost or status terms/labels, information grouping, and marking of important elements. Overall, these results indicate that Ulil Trans has not been consistently considered user-friendly by the majority of testers or users. As a result, the key ergonomic and usability improvements that must be addressed center on (1) enhancing the navigation and wayfinding design; (2) strengthening menu/icon marking as well as affordance and consistency of interaction; and (3) increasing the hierarchy and clarity of information on the History and Invoice screens, rather than focusing primarily on the data/form input flow.

E. Batch 5: Phase 8 to 9 (final phases)

Following Phases 6-7, which established fundamental usability results through SUS and task-based testing, Phase 8 broadened the analysis by converting those findings into data visualizations that made patterns of obstacles simpler to understand at a glance.

Phase 8 clarified where barriers were most frequently encountered (particularly in homepage navigation and discoverability for low-experience users) and which areas indicated remaining interaction understandability or reliability issues.

The data reveals that Ulil Trans's overall usability remains mediocre, with an average SUS score of 57.73 and a score dispersion indicating an uneven user experience. The sample findings demonstrate that testers had never used Ulil Trans, and the majority have only used it once or twice, emphasizing the importance of first-time discovery and navigation. Across the scenario results, Task 1 (Home Navigation and Accessing Notifications) presented the greatest challenge, with more than half of respondents completing the task or scenario but finding it difficult, and the distribution of times hovered around 2 minutes, indicating inefficient exploration rather than complete failure.

The History and Invoice scenarios demonstrated better overall completion and comprehension, but still left a significant portion of users struggling, consistent with visuals indicating information density, unclear status cues, and weak design hierarchy; furthermore, at least one instance of an impressable icon presented an interaction reliability issue that could disproportionately lower perceived usability of the app. The order form performed relatively better in terms of completion, but the observed failure rate remained significant for the core conversion flow. Overall, these figures support prioritizing improvements that reduce navigational ambiguity (clear global navigation and consistent back/home behavior), address interaction reliability (tap targets and responsive UI elements), and improve the information design on the History and Invoice pages through clearer labeling, a scannable layout, and prominent primary actions.

IV. CONCLUSION

By visualizing and calculating data, this research indicates that the Ulil Trans mobile app's performance is greatly determined by how well its interface and interaction design match the characteristics of its target users and the environment in which it is used. Differences in digital literacy, linguistic expectations, accessibility requirements, device limits, and time constraints affect how users understand interface signals, traverse displays, and perform activities in user-tested scenarios. When these contextual elements are not clearly handled, problems arise as recurrent impediments in the app's usage flow, such as discovering functionalities, comprehending labels and status information, properly inputting data, and identifying system feedback.

These findings also reinforce that iterative, user-centered design cycles are not a silver bullet, but rather a necessary tool for reducing barriers. Contextual inquiry and task analysis help clarify the intended use of an app and its specific constraints, while interactive prototyping and iterative usability testing reveal where interactions are struggling in practice. As interface clarity improves and interaction effort decreases, users are more likely to consistently complete tasks with fewer errors, contributing to higher acceptance, continued use, and more reliable achievement of app goals.

To address the identified limitations, this study presents a new interface approach that focuses on enhancing important user processes while also boosting clarity and feedback. Recommended interface improvements include: (1) restructuring the navigation hierarchy so that core features are more accessible; (2) improving information clarity through more explicit labels, consistent iconography, and easily understandable history summaries; (3) improvements to feedback and error prevention mechanisms (confirmation for risky message actions, clear progress indicators); and (4) adapting the interface to device accessibility and context. In this approach, the suggested interface is not only more visually appealing, but it is also intended to reduce cognitive burden, minimize user mistakes, and speed up job completion.

ACKNOWLEDGEMENT

Finally, this study has practical implications for design and evaluation: interface modifications should be prioritized based on friction points that have the largest influence on crucial usage and confirmed with performance-based usability indicators. Future study might enhance these findings by testing a larger range of user segments, assessing use in more natural circumstances, and tracking long-term usage to see if usability benefits transfer into maintained engagement over time.

REFERENCES

[1] F. Schewe and M. Vollrath, “Ecological interface design effectively reduces cognitive workload – The example of HMIs for speed control,” *Transportation Research Part F:*

Traffic Psychology and Behaviour, vol. 72, pp. 155–170, Jul. 2020, doi: 10.1016/j.trf.2020.05.009.

[2] Q. Jiang, L. Deng, J. Zhang, and Y. Pengbo, “User-Centered Design Strategies for Age-Friendly Mobile News Apps,” *Sage Open*, vol. 14, no. 4, p. 21582440241285393, Oct. 2024, doi: 10.1177/21582440241285393.

[3] A. Durmuş, “The influence of digital literacy on mHealth app usability: The mediating role of patient expertise,” *DIGITAL HEALTH*, vol. 10, p. 20552076241299061, Jan. 2024, doi: 10.1177/20552076241299061.

[4] J. Gulliksen, B. Göransson, I. Boivie, S. Blomkvist, J. Persson, and Å. Cajander, “Key principles for user-centred systems design,” *Behaviour & Information Technology*, vol. 22, no. 6, pp. 397–409, Nov. 2003, doi: 10.1080/01449290310001624329.

[5] H. Hoehle and V. Venkatesh, “Mobile Application Usability: Conceptualization and Instrument Development1,” *MIS Quarterly*, vol. 39, no. 2, pp. 435–472, Jun. 2015, doi: 10.25300/MISQ/2015/39.2.08.

[6] P. Weichbroth, “Usability Testing of Mobile Applications: A Methodological Framework,” *Applied Sciences*, vol. 14, no. 5, p. 1792, Feb. 2024, doi: 10.3390/app14051792.

[7] P. Weichbroth, “Usability of Mobile Applications: A Systematic Literature Study,” *IEEE Access*, vol. 8, pp. 55563–55577, 2020, doi: 10.1109/ACCESS.2020.2981892.

[8] T. F. Hasan, T. Wahyuningrum, and A. C. Wardhana, “USABILITY TESTING PADA M-COMMERCE MENGGUNAKAN KUESIONER USE (USEFULNESS, SATISFACTION, AND EASE OF USE) DAN PERFORMANCE TEST (STUDI KASUS : TOKOPEDIA)”.

[9] Z. Chen, X. Lan, J. Piao, Y. Zhang, and Y. Li, “A Mixed-Methods Analysis of the Algorithm-Mediated Labor of Online Food Deliverers in China,” *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, pp. 1–24, Nov. 2022, doi: 10.1145/3555585.

[10] B. A. Kumar, S. S. Chand, and M. S. Goundar, “Usability testing of mobile learning applications: a systematic mapping study,” *IJILT*, vol. 41, no. 2, pp. 113–129, Apr. 2024, doi: 10.1108/IJILT-03-2023-0029.

[11] A. Generosi, J. Y. Villafan, L. Giraldi, S. Ceccacci, and M. Mengoni, “A Test Management System to Support Remote Usability Assessment of Web Applications,” *Information*, vol. 13, no. 10, p. 505, Oct. 2022, doi: 10.3390/info13100505.

[12] T. L. Mardi Suryanto, A. Faroqi, and W. N. Simarmata, “SYSTEM USABILITY SCALE (SUS) SEBAGAI METODE PENGUJIAN KEGUNAAN PADA SITUS PROGRAM STUDI,” *sitasi*, vol. 2, no. 1, pp. 285–294, Sep. 2022, doi: 10.33005/sitasi.v2i1.314.

[13] L. Karlovska, A. Petrasova, V. Petras, and M. Landa, “Redesigning Graphical User Interface of Open-Source Geospatial Software in a Community-Driven Way: A Case Study of GRASS GIS,” *IJGI*, vol. 12, no. 9, p. 376, Sep. 2023, doi: 10.3390/ijgi12090376.

[14] N. K. S. Ardhani, N. K. A. Wirdiani, and N. K. D. Rusjayanthi, “Evaluation and Improvement of User Interface Design in Mobile Delivery Services With User Centered Design Approach and Heuristic Evaluation (Case Study: Kirimaja Application),” *JIM*, vol. 11, no. 3, p. 211, Dec. 2023, doi: 10.24843/JIM.2023.v11.i03.p07.

[15] A. Maravić, V. Pajić, and M. Andrejić, “A Holistic Human-Based Approach to Last-Mile Delivery: Stakeholder-Based Evaluation of Logistics Strategies,” *Logistics*, vol. 9, no. 4, p. 135, Sep. 2025, doi: 10.3390/logistics9040135.

[16] M. Hyzy *et al.*, “System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis,” *JMIR Mhealth Uhealth*, vol. 10, no. 8, p. e37290, Aug. 2022, doi: 10.2196/37290.

[17] T. M. Saddyah and S. P. Saragih, “PERANCANGAN UI/UX DELIVERY MOBILE APP DENGAN METODE DESIGN THINKING DAN USABILITY SCALE,” *CBIS*, vol. 12, no. 1, pp. 39–51, Mar. 2024, doi: 10.33884/cbis.v12i1.8242.

[18] R. F. Adler, K. Baez, P. Morales, J. Sotelo, D. Victorson, and S. Magasi, “Evaluating the Usability of an mHealth App for Empowering Cancer Survivors With Disabilities: Heuristic Evaluation and Usability Testing,” *JMIR Hum Factors*, vol. 11, p. e51522, Apr. 2024, doi: 10.2196/51522.

[19] I. Hussain, I. A. Khan, W. Jadoon, R. N. Jadoon, A. N. Khan, and M. Shafi, “Touch or click friendly: Towards adaptive user interfaces for complex applications,” *PLoS ONE*, vol. 19, no. 2, p. e0297056, Feb. 2024, doi: 10.1371/journal.pone.0297056.

[20] Z. Galavi, S. Norouzi, and R. Khajouei, “Heuristics used for evaluating the usability of mobile health applications: A systematic literature review,” *DIGITAL HEALTH*, vol. 10, p. 20552076241253539, Jan. 2024, doi: 10.1177/20552076241253539.

[21] K. Christianto, A. Chakir, J. F. Andry, F. Adikara, L. Liliana, and J. Felicia, “Modeling User Experience in Delivery Applications Using the Design Thinking Method and System Usability Scale,” *Journal of Computer Science*, vol. 20, no. 7, pp. 722–729, Jul. 2024, doi: 10.3844/jcssp.2024.722.729.

[22] E. D. Rinawiyanti, “Mixed Method Usability Testing for User Experience and User Interface of AI-Based Supermarket Applications,” *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 483–495, Jan. 2024, doi: 10.47738/jads.v6i1.453.

[23] A. Kaya, R. Ozturk, and C. Altin Gumussoy, “Usability Measurement of Mobile Applications with System Usability Scale (SUS),” in *Industrial Engineering in the Big Data Era*, F. Calisir, E. Cevikcan, and H. Camgoz Akdag, Eds., in Lecture Notes in Management and Industrial Engineering. , Cham: Springer International Publishing, 2019, pp. 389–400. doi: 10.1007/978-3-030-03317-0_32.

[24] R. Andika and D. Renaldi, “Heuristic Evaluation of UI/UX to Enhance Experience and Sales in E-commerce,” *bit-Tech*, vol. 7, no. 2, pp. 224–251, Dec. 2024, doi: 10.32877/bt.v7i2.1730.