Evaluating User Experience in a Public Sector Digital System Through Nielsen's Heuristic Approach

Nagyan Yosse Wibisono[1], Viany Utami Tjhin[2]
Department Information Systems Management, Jakarta, Indonesia[1]
BINUS Graduate Program—Master of Information System Management, Bina Nusantara University, Indonesia[2]

***Abstract*—The digital transformation in public administration has encouraged the Indonesian National Police (Polri) to adopt a digital government application for managing official documents electronically. Despite its functional benefits, users have reported several usability issues such as non-intuitive navigation, inconsistent interface design, inadequate system feedback, and insufficient documentation. To systematically address these problems, this study combines Jakob Nielsen's Heuristic Evaluation (HE) with Partial Least Squares Structural Equation Modeling (PLS-SEM), offering a hybrid methodological approach that is rarely applied in public sector UX evaluation. Data were collected through a structured questionnaire distributed to 156 active users of the application. The instrument measured ten heuristic principles and user experience dimensions using a four-point Likert scale. The results reveal that all heuristic dimensions scored within the "good" range with mean values between 3.19 and 3.38, classified as cosmetic issues under Nielsen's severity scale. More importantly, the analysis shows that only three heuristics—Match Between System and the Real World, Help Users Recognize Diagnose and Recover from Errors, and Help and Documentation—have a significant positive impact on user experience perceptions. Together, these heuristics explain 90.9 per cent of the variance in user experience, highlighting their critical role in shaping user-centered digital government systems. This study advances existing evaluation models by demonstrating the effectiveness of integrating heuristic evaluation with quantitative SEM-based analysis, bridging diagnostic insights with statistical rigor. The findings provide a prioritized roadmap for improving the application's interface and emphasize the importance of user-centered design for enhancing the adoption and effectiveness of public sector digital systems.**

Keywords*—Usability evaluation; heuristic evaluation; user experience; digital government; PLS-SEM; public sector technology; UX mining; usability modeling

I. Introduction

Digital transformation has become a critical agenda in modern public sector organizations, particularly in efforts to enhance efficiency, transparency, and service quality. The Indonesian National Police (Polri), as a key state institution, has also embraced this shift through the implementation of a digital government application aimed at managing internal administrative processes such as task letters and official memos. This application replaces conventional paper-based workflows with an integrated digital platform, allowing for faster document processing, electronic validation, and improved traceability of administrative activities. Despite the functional advantages offered by such systems, their adoption and long-term success depend heavily on usability and the overall user experience (UX). Prior studies indicate that inadequate usability can significantly hinder system acceptance, reduce user satisfaction, and increase operational errors in public service platform [1][2]. In the context of this digital application used by Polri personnel, several recurring issues were identified through preliminary user feedback: non-intuitive navigation, inconsistency across interface layouts, insufficient system feedback, and the absence of clear user documentation. These problems, although often overlooked during development, directly affect user efficiency and satisfaction, especially in mission-critical environments such as law enforcement institutions.

To regularly evaluate the system's usability and understand how it impacts the user experience, this study employs the Heuristic Evaluation (HE) framework introduced by Jakob Nielsen. The ten usability heuristics provide a comprehensive lens through which interface design can be assessed, identifying both minor and severe issues related to user interaction. To complement this qualitative framework with empirical rigor, this study integrates HE with Partial Least Squares Structural Equation Modeling (PLS-SEM), a powerful multivariate technique suitable for analyzing complex models involving latent variables. This combination allows researchers to both diagnose usability problems and quantify their influence on the perceived UX.

Existing literature has demonstrated the effectiveness of HE and PLS-SEM separately, yet few studies have attempted to integrate both methods in evaluating digital government applications. Moreover, limited research has explored this approach within the unique operational context of Indonesian law enforcement institutions. Therefore, this study addresses a significant research gap by applying a hybrid HE–PLS-SEM approach to assess the usability of a digital application used by Polri personnel. The outcome is expected to offer a robust model that explains the impact of each usability dimension on UX, along with severity-based prioritization of interface improvements.

The study not only contributes methodologically by combining heuristic diagnostics with structural modeling but also provides practical insights for public sector developers and decision-makers. By identifying which usability principles most strongly affect user experience, the findings can inform future design and policy decisions aimed at enhancing the effectiveness of digital governance tools.

Research Questions

Based on the identified research problem, this study is guided by the following research questions:

- What usability issues are experienced by users of the Indonesian National Police's digital document system (Astina) as identified through Nielsen's heuristic evaluation?
- Which heuristic principles have the most significant influence on user experience, as measured using PLS-SEM?
- How can the findings be used to provide practical recommendations and a prioritized roadmap for improving the system's usability and supporting digital transformation in public sector organizations?

II. Liteaturew Review

A. User Experience

User Experience (UX) is an analytical process that focuses on understanding how users interact with digital products or services, including their perceptions, emotions, and levels of satisfaction during usage. UX exploration aims to identify user needs, challenges, and expectations towards a system, serving as the foundation for designing more intuitive, efficient, and user-centered products [3].

This process involves identifying desired features, analyzing issues or obstacles encountered during usage, and evaluating user satisfaction across various system aspects. Gathering user feedback is a key component in optimizing product design, enhancing the overall experience, and ensuring that the system is more responsive and efficient in meeting user needs [4].

Improving UX offers several benefits, including increased user satisfaction and loyalty, refinement of products or services to better align with user expectations, and enhanced effectiveness and efficiency of system usage. Furthermore, a deep understanding of user needs can foster innovation and generate new ideas that are more relevant to the market, resulting in continuous improvements and added value for users [5].

B. User Interface

User Interface (UI) plays a vital role in shaping the overall customer experience, acting as the visual and interactive layer between users and digital systems. In computing environments, the three key components—hardware, software, and human users (brainware)—are interconnected through the UI, which serves as the bridge that enables effective communication and interaction [6].

UI is central to the implementation of user-centered design (UCD), a fundamental principle of UI/UX that emphasizes placing users at the core of the design process. This approach is especially critical in financial applications, where usability and trust heavily influence the user experience. UCD involves deeply understanding user needs, behaviors, and constraints to design interfaces that address those elements effectively. In financial contexts, this might include simplifying complex financial tasks, presenting information clearly, and ensuring intuitive navigation [7].

Functionally, UI comprises two main components: input and output. Input elements involve user commands delivered through devices like keyboards or mice, while output components deliver system responses via visuals, sounds, or other feedback mechanisms. An effectively designed UI enhances user experience, improves system efficiency, and ensures ease of access to and comprehension of information [8].

C. Heuristic Evaluation

UX heuristics are a set of principles used to evaluate user interface designs to ensure they are intuitive, efficient, and user-friendly. These guidelines are commonly applied in heuristic evaluations to identify usability issues without direct user testing The goal is to improve interface elements based on user needs and expectations, facilitating efficient task completion and enhancing overall user satisfaction [8].

Key heuristics include [9]:

- Visibility of system status – keeping users informed about system activities through timely feedback.
- Match between system and the real world – using familiar language and conventions to ensure intuitive interactions.
- User control and freedom – providing clearly marked exits to prevent user frustration from unintended actions.
- Consistency and standards – aligning with platform conventions to reduce cognitive load (Jakob's law).
- Error prevention – designing to prevent problems before they occur through validation and confirmation.
- Recognition rather than recall – minimizing memory load by keeping key information visible.
- Flexibility and efficiency of use – allowing expert users to use shortcuts while supporting novice users.
- Aesthetic and minimalist design – presenting only relevant information to enhance clarity and usability.
- Help users recognize, diagnose, and recover from errors – providing user-friendly messages and solutions.
- Help and documentation – offering accessible, task-oriented help when needed.

These heuristics contribute to designing systems that are not only functional but also provide a positive and satisfying user experience [10].

III. Objectives

The primary objective of this study is to evaluate the usability and user experience (UX) of a digital government application implemented by the Indonesian National Police (Polri) to support internal administrative processes. Specifically, the research aims to:

1) Determine the severity rating level of each usability problem found in the application, as a basis for prioritizing system improvements.

2) Analyze the influence of usability aspects on the user experience (UX) of users in using the Astina application.

3) Identify and categorize significant usability problems based on the perceptions of active users of the Astina application.

Through these objectives, this study aims to offer a validated usability assessment model that can be applied to other digital government systems, while also supporting the adoption of user-centered design principles in public sector digital transformation.

IV. Methods

A. Research Design

This study employed a quantitative research design using a survey-based approach to evaluate the usability and user experience (UX) of a digital government application used by the Indonesian National Police (Polri). The research integrated Heuristic Evaluation (HE) principles developed by Jakob Nielsen with Partial Least Squares Structural Equation Modeling (PLS-SEM) to statistically test the relationships between usability factors and user experience outcomes.

B. Instrument Development

The research instrument was a structured questionnaire consisting of 33 items grouped under 11 latent variables: 10 variables represented Nielsen's usability heuristics:

One variable captured user experience (UX), measured by three indicators: perceived ease of use, satisfaction, and efficiency. Each heuristic variable was operationalized using three indicators, making a total of 30 indicators for the usability construct. Responses were recorded using a 4-point Likert scale (1 = strongly disagree, 4 = strongly agree), selected to encourage decisive responses and avoid neutrality.

The instrument was validated by domain experts and tested in a pilot study involving 10 users. Subsequently, the internal consistency and construct validity were confirmed through Cronbach's Alpha and Pearson correlation analyses.

C. Sampling and Data Collection

The population consisted of active users of the digital application within Polri. Using purposive sampling, data were collected from 156 respondents who met the inclusion criteria:

Had used the application at least once for official administrative tasks; Were willing to participate and complete the survey fully. The survey was administered online using Google Forms, and data collection took place over two weeks in April 2025. This approach was practical for reaching respondents across various units within Polri and ensured secure and accessible participation.

D. Sampling and Data Collection

Severity Rating

Usability issues were categorized using Nielsen's severity rating scale (0–4), adapted for mean scores on the Likert scale:

0 = no usability problem (mean > 3.4)

1 = cosmetic problem

2 = minor usability problem

3 = major usability problem

4 = critical usability issue (mean < 1.59)

Each heuristic indicator's mean score was converted into a severity level to prioritize improvement areas.

Structural Equation Modeling (PLS-SEM)

PLS-SEM was conducted using SmartPLS v4 to assess both measurement and structural models:

- Outer model: Evaluated using convergent validity (outer loadings ≥ 0.7, AVE ≥ 0.5), discriminant validity (Fornell–Larcker), and reliability (Cronbach's Alpha ≥ 0.7).
- Inner model: Assessed using path coefficients, R^2 values, and bootstrapping (5,000 resamples) to determine the significance of relationships between heuristic variables and UX.

This two-stage analysis approach ensured that the model was both statistically sound and meaningful for theoretical and practical applications.

V. Result

A. Severity Rating

The average severity rating results for each heuristic dimension (H1–H10) are also in the same category, with a score value between 3.19 and 3.29. All dimensions received a severity rating of 1, which means that no minor, major, or critical problems were found in the main aspects of the application interface. A summary of the severity rating of all variables is presented in:

Several indicators that have scores close to the upper limit, such as HE12 (3.30) and HE23 (3.38) indicate very positive user perceptions of the elements, approaching a problem-free condition. Meanwhile, indicators with the lowest scores, such as HE102 (3.14) and HE103 (3.19) remain in the "good" category, but can be considered for quality improvement if you want to achieve an optimal user experience. The average severity rating results for each heuristic dimension (H1–H10) are also in the same category, with a score value between 3.19 and 3.29. All dimensions get a severity rating of 1, which means that no minor, major, or critical problems were found in the main aspects of the application interface. A summary of the severity ratings for all variables is presented in Tabel I.

B. PLS-SEM Analysis

The methodological procedure followed in this study included the distribution of structured questionnaires to 156 respondents, followed by data screening and coding. In the PLS-SEM analysis, the measurement (outer) model was first evaluated through convergent validity, discriminant validity, and reliability tests using indicators such as factor loadings, Average Variance Extracted (AVE), Composite Reliability (CR), and Cronbach's Alpha. Next, the structural (inner) model was assessed through path coefficients, the coefficient of determination (R Square), predictive relevance (Q Square), and effect size (f Square). Finally, bootstrapping with 5,000 resamples was employed to test the significance of the hypothesized relationships (Table I).

TABLE I. SEVERITY RATING

Code	Average Score	Quality	Severity	Interpretation
H1	3,29	Good	1	Cosmetic Problem
H2	3,29	Good	1	Cosmetic Problem
H3	3,27	Good	1	Cosmetic Problem
H4	3,24	Good	1	Cosmetic Problem
H5	3,26	Good	1	Cosmetic Problem
H6	3,29	Good	1	Cosmetic Problem
H7	3,20	Good	1	Cosmetic Problem
H8	3,22	Good	1	Cosmetic Problem
H9	3,20	Good	1	Cosmetic Problem
H10	3,19	Good	1	Cosmetic Problem

Partial Least Squares Structural Equation Modeling (PLS-SEM) is used in this study to analyze the relationship between Nielsen's 10 usability heuristic principles (HE1–HE10) and the perception of User Experience (UX) in Astina application users. This method was chosen because PLS-SEM is very suitable for predictive and complex models and can accommodate a relatively moderate number of samples with a strictly non-normal data distribution. PLS-SEM analysis is carried out through two main stages, namely: evaluation of the measurement model (outer model) and evaluation of the structural model (inner model). In addition, hypothesis testing is also carried out to determine the effect of each usability heuristic variable on UX.

Evaluation of the outer model aims to ensure that the indicators used can reflect the constructs being measured in a valid and reliable manner. The outer model test covers four main aspects: indicator reliability, construct reliability, convergent validity, and discriminant validity.

The reliability of an indicator in a reflective measurement model refers to the ability of an indicator to reflect a latent construct consistently and accurately. In PLS-SEM analysis, indicator reliability is indicated by the outer loading value of each item against its construct in Table II.

According to [11], an indicator is said to have adequate individual reliability if it has an outer loading value ≥ 0.70. A value above 0.70 indicates that more than 50% of the variance in the indicator can be explained by the latent construct. If the loading value is between 0.40–0.70, the indicator can be considered to be retained if the overall construct remains valid and reliable. However, indicators with loading values below 0.40 are generally recommended to be deleted.

In this study, the outer loadings results show that all indicators in the model have loading values between 0.800 and 0.905 (Table II) which means:

- All indicators pass the minimum threshold of 0.70, so they can be said to be very reliable.
- There is no indicator that is weak or needs to be eliminated from the model.
- The existence of each indicator contributes significantly to forming the relevant latent construct.

TABLE II. INDICATOR RELIABILITY (OUTER LOADING)

Indicator	Construct	Outer Loading	Result
HE11	HE1	0,865	Reliable
HE12	HE1	0,868	Reliable
HE13	HE1	0,846	Reliable
HE21	HE2	0,890	Reliable
HE22	HE2	0,800	Reliable
HE23	HE2	0,875	Reliable
HE31	HE3	0,874	Reliable
HE32	HE3	0,860	Reliable
HE33	HE3	0,853	Reliable
HE41	HE4	0,873	Reliable
HE42	HE4	0,872	Reliable
HE43	HE4	0,865	Reliable
HE51	HE5	0,831	Reliable
HE52	HE5	0,832	Reliable
HE53	HE5	0,853	Reliable
HE61	HE6	0,867	Reliable
HE62	HE6	0,896	Reliable
HE63	HE6	0,874	Reliable
HE71	HE7	0,882	Reliable
HE72	HE7	0,838	Reliable
HE73	HE7	0,868	Reliable
HE81	HE8	0,878	Reliable
HE82	HE8	0,881	Reliable
HE83	HE8	0,859	Reliable
HE91	HE9	0,880	Reliable
HE92	HE9	0,862	Reliable
HE93	HE9	0,881	Reliable
HE101	HE10	0,905	Reliable
HE102	HE10	0,876	Reliable
HE103	HE10	0,884	Reliable
UX1	UX	0,864	Reliable
UX2	UX	0,876	Reliable
UX3	UX	0,868	Reliable

These results indicate that each indicator functions well in measuring its respective construct. The very high correlation strength between the indicators and the latent construct provides a strong foundation that this data has high measurement quality, which is an important prerequisite before proceeding to the evaluation of the structural model and hypothesis testing.

Construct reliability is a measure to assess the internal consistency of the indicators that form a latent construct. This consistency shows how much the indicators together describe the construct they represent. In this study, construct reliability was evaluated using two main measures, namely Cronbach's Alpha and Composite Reliability (CR).

Cronbach's Alpha measures homogeneity between indicators and is a classic reliability indicator often used in quantitative research. A high alpha value indicates that the indicators are highly correlated with each other. According to (Hair Jr, 2021), a good Cronbach's Alpha value for both exploratory and confirmatory research is at least 0.60, while a value above 0.70 indicates adequate reliability.

Composite Reliability is considered a more accurate measure of construct reliability than Cronbach's Alpha, because it does not assume that all indicators have the same contribution to the construct. A good CR should have a minimum value of 0.70, and a value above 0.90 indicates very high reliability.

Based on the results of data processing using SmartPLS, the Cronbach's Alpha values ranged from 0.790 to 0.867, and the Composite Reliability values ranged from 0.877 to 0.918 for all constructs, as shown in Table III.

TABLE III. CONSTRUCT VALIDITY AND RELIABILITY

Construct	Cronbach's Alpha	Composite Reliability	AVE	Decision
HE1	0,823	0,895	0,739	Valid & Reliable
HE2	0,818	0,891	0,733	Valid & Reliable
HE3	0,828	0,897	0,744	Valid & Reliable
HE4	0,835	0,901	0,752	Valid & Reliable
HE5	0,790	0,877	0,704	Valid & Reliable
HE6	0,853	0,911	0,773	Valid & Reliable
HE7	0,829	0,898	0,745	Valid & Reliable
HE8	0,844	0,906	0,762	Valid & Reliable
HE9	0,847	0,907	0,764	Valid & Reliable
HE10	0,867	0,918	0,790	Valid & Reliable
UX	0,838	0,903	0,756	Valid & Reliable

All constructs in this model have Cronbach's Alpha values above 0.79 and Composite Reliability above 0.87, which means:

- Each latent construct is consistently measured by its indicators.
- No constructs need to be eliminated or revised because all have met the minimum construct reliability limit recommended in PLS-SEM research.
- High CR values (> 0.90) such as in HE10 and UX indicate that the indicators in the construct are very cohesive in measuring their latent concepts.

Thus, it can be concluded that all usability heuristic and user experience constructs in this model have very good internal consistency, and the results are suitable for use in the next structural analysis stage (inner model and hypothesis testing).

Convergent validity in PLS-SEM analysis is evaluated using the Average Variance Extracted (AVE) measure. According to, a construct is said to meet convergent validity if the AVE value is ≥ 0.50. This value indicates that at least 50% of the indicator variance can be explained by the latent construct it measures, while the rest comes from errors or other variables outside the model.

Based on the calculation results using SmartPLS, all constructs in the model — namely the 10 usability heuristic principles (HE1–HE10) and the User Experience (UX) variable — show very good AVE values, ranging from 0.704 to 0.790, as shown in Table III.

All constructs in this model meet the requirements for convergent validity because they have AVE values above 0.70, far exceeding the minimum threshold of 0.50. This shows that:

- Most of the variance of the indicators can be explained by the construct, not by measurement error or other variables.
- The indicators used have high correlation with each other and are cohesive in measuring the intended latent construct.
- The validity of the measurement model (outer model) as a whole is strong and can be relied on to continue to the evaluation of the relationship between constructs (inner model).

Thus, it can be concluded that all constructs in this research model have very good convergent validity, which is an important prerequisite to ensure that the results of the structural analysis (path analysis and hypothesis testing) are based on a valid measurement model.

Discriminant Validity in PLS-SEM analysis, discriminant validity is evaluated using the Fornell-Larcker Criterion, which compares the AVE square root value of each construct with its correlation value to other constructs. According to Fornell and Larcker (1981), a construct is said to have good discriminant validity if: The AVE square root value of a construct is greater than the correlation value between the construct and other constructs in the model.

Based on the results of data processing through SmartPLS, all constructs (HE1 to HE10 and UX) show AVE square root values (on the table diagonal) that are higher than the horizontal and vertical correlation values between other constructs, as shown in Table IV.

These results show that for each construct:

- The √AVE value (on the diagonal) is higher than the correlation value with other constructs (off the diagonal).
- This means that each construct is more closely related to its own indicators than to other constructs.

- This meets the Fornell-Larcker criteria and proves that the model has good discriminant validity.

Thus, it can be concluded that all constructs in this model are unique and empirically different from each other, and can be used independently in structural modeling.

TABLE IV. FORNELL LACKER CRITERION

Construct	HE1	HE2	HE3	HE4	HE5	HE6	HE7	HE8	HE9	HE10	UX
HE1	0,859										
HE2	0,782	0,856									
HE3	0,722	0,771	0,863								
HE4	0,752	0,833	0,746	0,867							
HE5	0,767	0,819	0,789	0,770	0,839						
HE6	0,732	0,801	0,755	0,816	0,810	0,876					
HE7	0,767	0,785	0,812	0,812	0,817	0,811	0,863				
HE8	0,702	0,796	0,756	0,775	0,795	0,844	0,836	0,873			
HE9	0,763	0,799	0,816	0,772	0,812	0,788	0,861	0,822	0,874		
HE10	0,750	0,792	0,755	0,809	0,830	0,805	0,824	0,823	0,840	0,889	
UX	0,778	0,883	0,801	0,827	0,816	0,808	0,844	0,811	0,898	0,887	0,869

Thus, all indicators and constructs in this model are declared valid and reliable, making them suitable for use in the evaluation of subsequent structural models.

Structural Model Evaluation (Inner Model**).** The evaluation of the inner model aims to test the causal relationship between exogenous constructs (HE1–HE10) and endogenous constructs (UX), as well as to measure how much the exogenous constructs contribute to explaining the UX variable.

Coefficient of Determination (R^2). The coefficient of determination (R^2) is the main indicator in the evaluation of the structural model (inner model) used to assess the predictive ability of exogenous constructs against endogenous constructs in the research model. In the context of this study, the endogenous construct in question is User Experience (UX), while the exogenous constructs are the ten usability heuristic principles (HE1 to HE10) based on the Nielsen framework (1994).

The R^2 value describes the proportion of variance from the endogenous variable (UX) that can be explained simultaneously by the exogenous variables in the model. The higher the R^2 value, the greater the ability of the exogenous construct to explain the endogenous construct.

According to Hair et al. (2021), the interpretation of the R^2 value in the context of the PLS-SEM model is as follows:

$R^2 \approx 0.75$ → substantial (strong)

$R^2 \approx 0.50$ → moderate (moderate)

$R^2 \approx 0.25$ → weak (weak)

The calculation results using SmartPLS show that the R Square (R^2) value for the User Experience (UX) variable is 0.909, with an adjusted R^2 of 0.902. These results are presented in the following Table V.

TABLE V. COEFFICIENT OF DETERMINATION R^2 AND ADJUSTED R^2

Endogenous Construct	R^2	Adjusted R^2	Interpretation
UX	0,909	0,902	Substantial

The R^2 value of 0.909 indicates that 90.9% of the variance in User Experience (UX) perception can be explained by the ten usability heuristic principles used in this model (HE1 to HE10). The remaining 9.1% is explained by other factors outside the model, such as user personal characteristics, infrastructure quality, or organizational context that are not included in the research construct.

This value is very high and falls into the substantial category, so it can be concluded that this model has very strong predictive ability. In other words, the usability heuristic principles formulated by Nielsen have been empirically proven to be able to explain UX perceptions in the use of the Astina application as a whole.

The reliability of R^2 is also strengthened by the adjusted R^2 of 0.902, which corrects the R^2 value based on the number of predictor variables in the model. This shows that the prediction results do not experience overfitting, and remain stable even considering the complexity of the model.

These findings are a strong indication that the usability heuristic model can be used as an effective theoretical framework in explaining and predicting the user experience of internal organizational applications. The high coefficient of determination also strengthens the argument that the user-centered approach in interface design is very relevant and has a significant impact on user perceptions of digital systems.

Path Coefficient and Hypothesis Testing. After the measurement model is declared valid and reliable, the next stage is to evaluate the relationship between latent constructs in the structural model (inner model) through path coefficient analysis and hypothesis testing. This analysis aims to test the significance of the influence of each exogenous construct, namely the ten usability heuristic principles (HE1–HE10), on the endogenous construct, namely user experience (UX).

Each hypothesis is tested by looking at the values:

Path Coefficient (β): shows the direction and magnitude of the influence.

- t-statistic: measures the statistical significance of the influence
- p-value: determines whether the influence is significant ($p < 0.05$)
- Decision: whether the hypothesis is accepted or rejected

Hypothesis testing was carried out using a bootstrapping procedure of 5,000 subsamples, as recommended by Hair et al. (2021), to obtain a sampling distribution and estimate the significance of the path. The results of the path coefficient analysis and hypothesis testing are shown in Table VI.

TABLE VI. PATH COEFFICIENT ANALYSIS AND HYPOTHESIS TEST RESULT

Hypothesis	Path	Path Coefficient (β)	t-statistic	p-value	Decision
H1	HE1 → UX	-0,009	0,112	0.911	Rejeted
H2	HE2 → UX	0,378	6,658	0,000	**Accepted**
H3	HE3 → UX	0,027	0,458	0,647	Rejeted
H4	HE4 → UX	0,029	0,536	0,592	Rejeted
H5	HE5 → UX	-0,091	1,560	0,199	Rejeted
H6	HE6 → UX	0,010	1,65	0,869	Rejeted
H7	HE7 → UX	0,058	0,853	0,394	Rejeted
H8	HE8 → UX	-0,082	1,402	0,161	Rejeted
H9	HE9 → UX	0,361	5,760	0,000	**Accepted**
H10	HE10 → UX	0,334	6,133	0,000	**Accepted**

Of the ten hypotheses proposed, only three were accepted, namely:

- H2 (HE2 → UX). The path coefficient of 0.378 with p = 0.000, indicates that the principle of "Match between system and the real world" has a positive and significant influence on UX. This means that a system that uses terms and logic that are in accordance with the user's way of thinking will improve the perception of user experience.
- H9 (HE9 → UX). The path coefficient of 0.361 and p = 0.000, indicates that a system that helps users recognize, diagnose, and recover errors effectively has a strong influence on user experience.
- H10 (HE10 → UX). With a coefficient of 0.334 and p = 0.000, these results indicate that the existence of easily accessible documentation and assistance also contributes significantly to improving UX.

Meanwhile, the other seven principles (HE1, HE3–HE8) show coefficient values that are not statistically significant ($p > 0.05$), so the hypothesis related to these principles is rejected. This indicates that in the context of using the Astina application, these principles do not have a strong direct influence on the perception of user experience.

Effect Size (f^2). In addition to evaluating the significance of the influence through path coefficients and hypothesis testing, the PLS-SEM structural model analysis also needs to consider the magnitude of the contribution of each exogenous construct to the endogenous construct, which is measured through the effect size. In this context, effect size is used to determine how much each usability heuristic principle (HE1–HE10) contributes to explaining the User Experience (UX) variable.

The interpretation of the f^2 value according to Cohen (1988) is as follows:

$f^2 \geq 0.35$ → large effect

$0.15 \leq f^2 < 0.35$ → medium effect

$0.02 \leq f^2 < 0.15$ → small effect

$f^2 < 0.02$ → not significant

The results of the calculation of the effect size (f^2) of each construct on UX are presented in Table VII.

TABLE VII. RESULTS OF THE CALCULATION OF THE EFFECT SIZE (F^2) OF EACH CONSTRUCT ON UX

Exogenous Construct	f^2	Effect Size Category
HE1 (Visibility of system status)	0,000	Not significant
HE2 (Match between system and the real world)	0,306	**Medium**
HE3 (User control and freedom)	0,002	Not significant
HE4 (Consistency and standards)	0,002	Not significant
HE5 (Error prevention)	0,018	Small
HE6 (Recognition rather than recall)	0,000	Not significant
HE7 (Flexibility and efficiency of use)	0,006	Not significant
HE8 (Aesthetic and minimalist design)	0,014	Small
HE9 (Help users recognize, diagnose, and recover from errors)	0,247	**Medium**
HE10 (Help and documentation)	0,232	**Medium**

From the table above, it can be seen that only three constructs have a medium effect size, namely:

- HE2 (Match between system and the real world) with $f^2 = 0.306$.
- HE9 (Help users recognize, diagnose, and recover from errors) with $f^2 = 0.247$.
- HE10 (Help and documentation) with $f^2 = 0.232$.

The three principles are not only statistically significant, but also have a fairly strong contribution in explaining UX, so they can be considered as key factors in the perception of user experience of the Astina application.

In contrast, constructs HE5 and HE8 show small effect sizes (f^2 <0.02–0.015), while HE1, HE3, HE4, HE6, and HE7 have very low or insignificant effect sizes, indicating that removing these constructs from the model will not have a significant impact on the R^2 value of UX.

The effect size evaluation strengthens the findings from the hypothesis test, that only a portion of Nielsen's usability heuristic principles provide a significant contribution to the perception of the UX of application users. This finding emphasizes that in the context of organizational or public sector applications such as Astina, the aspects of system communication, error support, and documentation are the interface design elements that have the most impact on users. Thus, the development of similar systems should focus on improving the quality of these aspects so that the user experience can be optimally improved.

Predictive Relevance (Q^2). In addition to evaluating the coefficient of determination (R^2) and effect size (f^2), the inner model analysis in the PLS-SEM approach also includes testing predictive relevance (Q^2). This test is used to determine whether the structural model has better predictive ability for endogenous constructs compared to models without exogenous constructs. The Q^2 test is carried out using the blindfolding technique, a procedure that systematically removes some data and then estimates the missing data based on the model. The Q^2 value obtained will indicate how well the model predicts the observed values of endogenous variables. Interpretation of the Q^2 value according to Hair et al. (2021) are as follows:

$Q^2 > 0.35 \rightarrow$ large predictive ability

$0.15 < Q^2 \leq 0.35 \rightarrow$ moderate predictive ability

$0 < Q^2 \leq 0.15 \rightarrow$ small predictive ability

$Q^2 \leq 0 \rightarrow$ no prediction (bad model)

Based on the blindfolding results in SmartPLS, the Q^2 value for the endogenous construct User Experience (UX) is 0.656, as shown in Table VIII.

TABLE VIII. PREDICTIVE RELEVANCE Q^2

Exogenous Construct	SSO	SSE	Q^2	Interpretation
UX	468.000	161.057	0,656	High predictive

The Q^2 value of 0.656 indicates that the model has a very strong predictive capacity for the User Experience (UX) construct. This means that the exogenous variables used (HE1 to HE10) are collectively able to predict UX with a high level of accuracy, far above the minimum threshold set in the PLS-SEM methodology ($Q^2 > 0.35$).

Thus, this research model is not only statistically significant (through p-value) and has a meaningful variable contribution (through f^2), but is also predictively relevant in the context of real users of the Astina application.

This finding strengthens the validity of the model as a whole, that the relationship between usability heuristics and UX built in this study is not only structurally valid, but also predictively applicable. The high Q^2 value confirms that the principles of usability heuristics chosen are indeed appropriate for use in developing or evaluating similar information systems, because the model is able to predict well how users will assess their experience when using the application. Thus, this model can be used as a reliable measuring tool for developing interface design and evaluating UX in information systems based on the organization's internal needs

Path Model Visualization. Fig. 1 below is a visual representation of the relationship between constructs in the structural model that has been analyzed using PLS-SEM. This model displays ten usability heuristic constructs (HE1 to HE10) as exogenous variables that lead to the user experience (UX) construct as an endogenous variable. In the model above, it can be seen that each exogenous construct (HE1 to HE10) has one arrow direction towards UX. The thickness of the arrow (if displayed) and the number above the arrow indicate the magnitude of the influence (path coefficient) of each variable on UX. Based on the path analysis and the results of the previous hypothesis test, it can be identified that three constructs, namely: HE2 (Match between system and the real world), HE9 (Help users recognize, diagnose, and recover from errors), HE10 (Help and documentation) have a significant and positive influence on UX, as indicated by the high path coefficient value and p-value <0.05. Meanwhile, the other seven constructs (HE1, HE3–HE8) showed no significant influence, with small path coefficient values and p-values > 0.05, which are depicted in the model as arrows with values close to zero. This model illustrates that not all usability heuristic principles have the same influence on the perception of user experience in the context of internal applications such as Astina. Only principles related to effective system communication, error support, and availability of documentation significantly affect UX. The visualization of this model strengthens the interpretation of the previous quantitative results, and provides an intuitive picture that a system design approach that focuses on communication, help, and guidance greatly determines user perceptions of the quality of their interactions with the system.

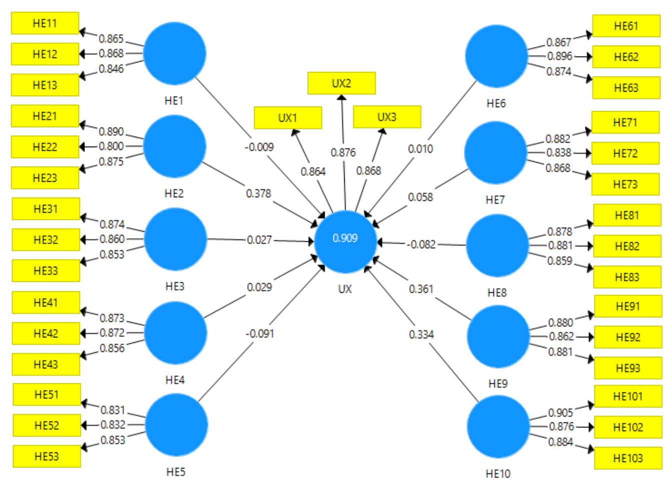

Fig. 1. PLS-SEM usability heuristic path model on UX.

VI. DISCUSSION

The results of this study aim to explain the influence of the usability heuristic principles developed by Nielsen (1994) on the perception of User Experience (UX) of Astina application users, which is a digital-based internal information system in the public sector. Testing was carried out using the PLS-SEM approach which allows simultaneous measurement between construct relationships and the predictive power of the model. Three

Heuristic Principles with Significant Influence on UX are discussed below.

A. Match Between System and the Real World

The HE2 construct has the highest path coefficient ($\beta = 0.378, p < 0.001$), indicating that the system's fit with the user's language, logical structure, and metaphor plays a major role in shaping UX perceptions. This principle emphasizes the importance of an interface that "speaks the user's language," as well as aligning interface elements with the user's real world to make it easier to understand (Nielsen, 1994).

This finding is supported who stated that user-language alignment is a key factor in reducing users' cognitive load and accelerating the adaptation of new systems. In the context of a public information system like Astina, where users come from a non-technical background, the use of familiar terms, intuitive icons, and organizational task-based navigation greatly helps improve the perception of experience. Research by [12] also showed that the difference between the technical terms of the system and the user's understanding can create interaction barriers, so that language alignment becomes essential in administrative systems.

Although HE2 has a statistically significant effect on UX, the severity rating results show an average score of 3.29, including the "Good" category, but it has not reached the "Very Good" category. This shows that users appreciate the conformity between the appearance and flow of the Astina system with the language and structure they understand; however, there are still minor shortcomings - such as the use of technical terms or symbols that are not yet fully intuitive - which cause them to rate the quality of the system as quite good but not yet ideal.

Implicitly, this means that users feel that this system is quite "familiar", but does not consistently reflect the user's everyday workflow or language. This reinforces the urgency of semantic and visual alignment between the system and the user's real world.

B. Help User Recognize, Diagnose, and Recover from Errors

The HE9 construct is also significant ($\beta = 0.361, p < 0.001$), indicating that the system's ability to handle user errors intelligently and informatively has a major impact on user experience. Users feel more confident and secure when the system is able to provide clear error information, as well as offer solutions or corrective steps.

System errors that cannot be handled directly by users are a major cause of frustration, and systems that fail to provide error feedback will hinder workflow. Meanwhile, Madzlan & Ishak (2024) added that good error recovery features (such as undo, warning dialogue, or error suggestion) are an integral component in building a resilient interface.

In a system like Astina, which is used in managing organizational administrative tasks, the existence of an error recovery system supports user efficiency and accuracy, and reduces the likelihood of fatal errors in the data input process.

HE9 also contributes significantly to UX, with an average severity rating score of the indicator between 3.16–3.22, which is included in the Good category. However, this score is one of the lowest among the other ten heuristic principles. This shows that although users feel the system helps them understand the errors that occur, there is still room for improvement, such as: clearer and more instructive error messages, automatic solutions or help that appears when an error occurs, undo or confirmation features before permanent action, Thus, HE9 is a critical element that is perceived as important, but has not been fully optimized in its implementation, as also emphasized by Azizi et al. (2021) in the health information system.

C. Help and Documentation

The HE10 construct also has a significant effect ($\beta = 0.334, p < 0.001$). Although this principle is usually considered a last resort, in the context of organizational applications with high complexity and moderate frequency of interaction, system documentation and assistance are still very much needed. Support in the form of tooltips, user guides, or interactive help systems provides a sense of security to new and old users.

Stated that accessible support systems not only help users solve problems but also foster the perception that the system is designed with empathy toward the user. This is also reinforced which shows that the availability of good documentation increases efficiency and perceived usability in hospital systems.

HE10 has a significant effect on UX and the average indicator score is 3.14–3.24, which is also included in the Good category, but is in the lowest range among other constructs. This means that even though help and documentation are considered important by users, current systems: may not provide enough contextual help, documentation is difficult to find or not available in the main interface, there is no searchable help or interactive guidance feature. This indicates a gap between expectations and implementation, where users consider that the existing help is limited or not easily accessible. This finding is in line with the study, which suggests that documentation must be on-demand, contextually relevant, and easy to navigate in order to truly contribute to UX.

Non-Significant Principles. The other seven principles (HE1, HE3–HE8) do not have a significant effect on the model. Although conceptually important, in the context of the Astina application, HE1 (Visibility of System Status) may be considered fulfilled, so that it is not a concern for users. Likewise, HE3 (User Control and Freedom), HE5 (Error Prevention), and HE8 (Aesthetic and Minimalist Design) may have less influence because the system is already quite standardized. In addition, HE6 (Recognition Rather Than Recall) and HE7 (Flexibility and Efficiency) may not have been optimized in terms of interface, so they have not created a direct effect on UX.

Research by [13] also found that not all usability heuristics have a direct influence on UX, depending on the application context and user characteristics. This means that theoretically important principles may have a statistically low influence if they are considered "default" by users.

Although only three of Nielsen's ten usability heuristic principles (HE2, HE9, and HE10) were proven to have a statistically significant effect on UX in the structural model, the other principles cannot be ignored. This is because the severity rating results show that all principles get an average score above

3.15, or fall into the "Good" category, with a severity classification at level 1 (cosmetic problems).

HE1 is not significant to UX ($\beta = -0.009$, $p = 0.911$), but gets a severity rating score of 3.29, indicating Good quality. This may be because the Astina system has provided quite good status indicators, such as notifications, loading status, or system messages, so that users are not too aware of their function explicitly because they are considered standard features. In other words, the absence of pain is not always the presence of pleasure — users do not feel disturbed, but also do not feel significantly helped.

HE3 is also not significant ($\beta = 0.027$, $p = 0.647$), but gets a rating score of 3.27. This shows that users value the freedom to navigate the system, for example through the back button, cancel, or menu selection. However, because the frequency of critical events that require this control may be low, users do not feel that the presence of this feature has a major impact on their experience.

HE4 obtained a small coefficient ($\beta = 0.029$, $p = 0.592$), with a severity score of 3.24, still in the "Good" category. This result indicates that the system has been quite consistent visually and functionally, but users may be less aware of minor inconsistencies, or insensitive to design standards, so that although theoretically important, it is not perceived as a dominant UX factor in the context of this application.

HE5 recorded a negative coefficient ($\beta = -0.091$, $p = 0.119$), with a severity score of 3.26. This is interesting, because even though users felt the system was good at preventing errors, they still valued the system more for providing solutions when errors occurred (HE9). In other words, error handling was perceived more than error prevention, especially in systems that were not very complex to use.

HE6 was also insignificant ($\beta = 0.010$, $p = 0.869$), with a severity score of 3.29. This suggests that users value interfaces that help them access information without having to remember a lot of things, such as dropdowns, labels, and icons. However, because these features run "behind the scenes," their contribution to UX is not considered dominant in their overall assessment.

HE7 had a low path coefficient ($\beta = 0.058$, $p = 0.394$), and a severity score of 3.20. In a system like Astina that is administrative and standardized, flexibility and efficiency may not be the main needs of users, so even though the quality is good, its impact on the overall UX is not significant. HE8 recorded a negative coefficient ($\beta = -0.082$, $p = 0.161$), but the severity score of 3.22, remains at the "Good" level. A clean and uncomplicated system does increase user comfort, but in a system with dominant functional functions, aesthetic appearance has not been a major factor in user satisfaction or comfort [14]. This explains why the perception remains positive, but its impact on UX is not statistically significant.

VII. Conclusion and Contribution

This study contributes methodologically by combining Nielsen's Heuristic Evaluation, which is widely used for diagnosing usability issues qualitatively, with Partial Least Squares Structural Equation Modeling, which provides quantitative validation of causal relationships. Compared to traditional heuristic evaluation that mainly identifies problems without measuring their statistical significance, the hybrid approach allows for both problem identification and empirical testing of their influence on user experience. In contrast to other UX evaluation models such as the System Usability Scale (SUS), User Experience Questionnaire (UEQ), or the Technology Acceptance Model (TAM), which focus on overall user perceptions, the HE–PLS-SEM integration offers a more detailed diagnostic framework. It not only highlights which usability heuristics matter most but also quantifies their contribution to UX outcomes, thereby advancing existing evaluation methods in both rigor and practical applicability for public sector systems.

The originality of this study lies in the integration of heuristic evaluation with PLS-SEM in evaluating a government application. While both methods have been widely applied independently in prior UX studies, their combination in the context of public sector systems, particularly the Indonesian National Police application, has not been previously explored. Prior research using heuristic evaluation has mainly produced qualitative insights into usability issues without quantifying their statistical significance, whereas studies applying PLS-SEM have typically examined user perceptions using generic acceptance models such as TAM or UTAUT without direct diagnostic links to usability heuristics. By merging these two approaches, this study not only identifies usability problems but also validates their causal impact on user experience outcomes. This represents a novel contribution, offering both methodological advancement and practical guidance for improving digital government systems.

Data Availability

Dataset is available from the Zenodo Repository, DOI: 10.5281/zenodo.16890311.

CRediT authorship contribution statement

Nagyan Yosse Wibisono: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft. Viany Utami Tjhin Supervision, Project administration, Funding acquisition, Writing – Review & Editing.

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] Azizi, A., Maniati, M., Ghanbari-Adivi, H., Aghajari, Z., Hashemi, S., Hajipoor, B., Qolami, A. R., Qolami, M., & Azizi, A. (2021). Usability evaluation of hospital information system according to heuristic evaluation. *Frontiers in Health Informatics*, *10*. doi: 10.30699/fhi.v10i1.271

[2] Galavi, Z., & Khajouei, R. (2023). Proposing an initial model for the heuristic evaluation of mHealth applications. *Informatics in Medicine Unlocked*, *41*. doi: 10.1016/j.imu.2023.101322

[3] Pei, X., Fu, S., & Jiang, T. (2023). An empirical study on user experience evaluation of VR interface in digital museums. *Data and Information Management*, *7*(4). doi: 10.1016/j.dim.2023.100057

[4] Lehnert, F., Doublet, S., & Sim, G. (2025). Designing usability/user experience heuristics to evaluate e-assessments administered to children. *Computer Standards & Interfaces*, *92*, 103933. doi: 10.1016/J.CSI.2024.103933

[5] Macias-Velasquez, S., Medellin-Castillo, H. I., & Garcia-Barrientos, A. (2024). New-user experience evaluation in a semi-immersive and haptic-enabled virtual reality system for assembly operations. *International Journal of Human-Computer Studies*, *190*, 103312. doi: 10.1016/J.IJHCS.2024.103312

[6] Faudzi, M. A., Cob, Z. C., Ghazali, M., Omar, R., & Sharudin, S. A. (2024). User interface design in mobile learning applications: Developing and evaluating a questionnaire for measuring learners' extraneous cognitive load. *Heliyon*, *10*(18), e37494. doi: 10.1016/J.HELIYON.2024.E37494

[7] Quiñones, D., Ojeda, C., Herrera, R. F., & Rojas, L. F. (2024). UXH-GEDAPP: A set of user experience heuristics for evaluating generative design applications. *Information and Software Technology*, *168*, 107408. doi: 10.1016/J.INFSOF.2024.107408

[8] International Standard. (2019). Ergonomics of human-system interaction-Human-centred design for interactive systems COPYRIGHT PROTECTED DOCUMENT. Retrieved from www.iso.org

[9] Nielsen Norman Group (2025) 10 Usability Heuristics for User Interface Design

[10] Pierre, R. da S. de Q. (2015). Heuristics in Design: A Literature Review. *Procedia Manufacturing*, *3*, 6571–6578. doi: 10.1016/J.PROMFG.2015.07.961

[11] K. Eli Hair Jr. (2021). Classroom Companion: Business Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R AAWorkbook.ssa, "Title of paper if known," unpublished.

[12] Bouraghi, H., Rezayi, S., Amirazodi, S., Nabovati, E., & Saeedi, S. (2022). Evaluating the usability of a national health information system with heuristic method. *Journal of Education and Health Promotion*, *11*(1), 182. doi: 10.4103/jehp.jehp_349_21

[13] Andika, R., & Renaldi, D. (2024). Heuristic Evaluation of UI/UX to Enhance Experience and Sales in E-commerce. *Bit-Tech*, *7*(2), 224–251. doi: 10.32877/bt.v7i2.1730

[14] Fatkhuri, Dharmawan, D., Desty Febrian, W., Karyadi, S., & Sani, I. (2024). Application of Heuristic Evaluation Method to Evaluate User Experience and User Interface of Personnel Management Information Systems to Improve Employee Performance. *Jurnal Informasi Dan Teknologi*, 14–20. doi: 10.60083/jidt.v6i1.466