Usability And Cognitive Load Evaluation Of The Threads Interface Using The System Usability Scale (SUS) And NASA-TLX Methods

Kifatul Hasna Tria Widyarsi[1*], Yayak Kartika Sari[2], Taufiq Agung Cahyono[3]

[1,2,3]Fakultas Sains dan Teknologi, Universitas Bhinneka PGRI Tulungagung, Indonesia
[*] Corresponding author:
Email: [1*]kifatulhasna123@gmail.com

Abstract.
The rapid evolution of social media interfaces has reshaped how users consume digital information, and interface complexity is known to affect both ease of use and the mental effort users must invest. This study examines the level of usability and cognitive load of the Threads interface and the relationship between the two constructs, using a quantitative descriptive-correlational approach. An online questionnaire was distributed to 200 students of Universitas Bhinneka PGRI Tulungagung selected through purposive sampling. Usability was measured with the System Usability Scale (SUS) and cognitive load with the NASA Task Load Index (NASA-TLX); data were processed in IBM SPSS. All instruments were valid and reliable, with Cronbach's Alpha values of 0.728 (SUS) and 0.786 (NASA-TLX). The mean SUS score was 70.0875, placing Threads in the "Good" category, while the mean NASA-TLX score of 58.83 indicated a moderate level of cognitive load. A Pearson correlation coefficient of 0.474 (p = 0.000) and a simple linear regression equation of Y = 0.735 + 0.521X, with an R Square of 0.225, showed that usability contributed 22.5% to cognitive load. The findings indicate that Threads is reasonably user-friendly but still imposes a moderate cognitive burden on its users, and that this burden rises together with—rather than in opposition to—perceived usability.

Keywords: *Usability, Cognitive Load, System Usability Scale (SUS), NASA-TLX and Threads.*

I. INTRODUCTION

Social media platforms have fundamentally altered the way people consume and interact with digital information. Poorly designed interfaces increase cognitive load, reduce engagement, and lower user retention [1], while the overall usability of a system strongly shapes user satisfaction and the adoption of modern digital applications [2]. As new platforms such as Threads combine text-based interaction with cross-platform integration, the complexity of their features and connections can heighten the mental workload required of users, making it harder for them to interpret content, follow commands, and complete tasks effectively. Usability and cognitive load are therefore both critical dimensions to examine when studying how people engage with contemporary digital applications.

Threads has grown rapidly since its 2023 launch. By mid-2025 the platform had surpassed 400 million monthly active users, and its mobile application recorded 115.1 million daily active users in June 2025 — a 127.8% year-over-year increase that brought it close to the daily mobile usage of its main competitor, X [11]. This growth has attracted academic interest as a digital social phenomenon worth investigating, including comparative studies of user sentiment between Threads and X that reveal differing user perceptions across the two platforms [8].

In Indonesia, the launch of Threads was met with considerable enthusiasm because of its integration with Instagram, which allows new users to instantly connect with an existing social network. However, several local studies suggest that the user experience is not yet fully satisfactory; the integration with Instagram has, in some cases, added complexity for users trying to manage accounts across both platforms, potentially raising mental load [8]. Local research on Threads has also found that user experience dimensions such as clarity, efficiency, and dependability receive only moderate ratings, suggesting that the user experience still has room for improvement [8].

A preliminary observation involving five students at Universitas Bhinneka PGRI found that students tend to voice opinions, academic complaints, and personal views more actively on Threads than on other platforms. This observation suggests that platform characteristics can shape the cognitive load users experience: while Threads' minimalist, text-focused design may offer cognitive ease relative to more visually

dense platforms, some students still reported confusion around specific feature flows—such as privacy settings and interactions between posts—as well as fatigue caused by repeated notifications and continuously refreshing content.

The System Usability Scale (SUS) is one of the most widely used instruments for evaluating usability in both academic and industrial settings, owing to its simplicity, brief administration time, and ability to produce a reasonably accurate assessment of an application's usability quality [1]. SUS classifies scores into categories ranging from Not Acceptable to Excellent, helping researchers and developers understand how comfortable an application is to use [12].

Beyond usability, cognitive load is another critical factor in evaluating the quality of a digital interface, since users may experience mental fatigue, loss of focus, and task difficulty when cognitive demands are excessive. The NASA Task Load Index (NASA-TLX) is the most widely used instrument in Indonesian research for measuring cognitive load across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration level [2]. NASA-TLX has proven effective at providing a comprehensive picture of mental workload across various application contexts, including digital transportation and public service applications, making it well-suited for assessing the cognitive load users invest when using Threads.

Numerous studies conducted in Indonesia have examined the relationship between usability and cognitive load, showing that the two factors are interdependent: when an interface's design is suboptimal, users must exert greater mental effort to understand navigation or access specific features, thereby increasing cognitive load [7]. Conversely, even an application rich in well-functioning features can suffer declining user perception if the mental burden it imposes is too great. Issues specific to Threads—such as its distinctive conversation structure, cross-platform integration, and frequent interface changes during updates—all add to users' mental burden. A thorough evaluation using SUS and NASA-TLX can therefore provide deeper insight into the quality of the Threads interface and the extent of cognitive load its users experience.

Other evaluation methods, such as Heuristic Evaluation, Cognitive Walkthrough, and the USE Questionnaire, were also considered but present limitations for this context. Heuristic Evaluation relies heavily on expert judgment rather than actual user experience; Cognitive Walkthrough is well suited to structured tasks but inadequate for predicting user experience in social/text-based media such as Threads; and the USE Questionnaire can produce less accurate data when respondents experience fatigue from answering a large number of items.

Based on this background, this study evaluates the usability and cognitive load of the Threads interface using SUS and NASA-TLX. Specifically, it seeks to answer: (1) what is the level of usability of the Threads interface as assessed by users through SUS, and (2) what is the level of cognitive load experienced by users interacting with Threads as measured through NASA-TLX. The study is expected to enrich the academic literature on Human-Computer Interaction (HCI) and User Experience (UX) evaluation of contemporary social media, while offering developers concrete direction for improving the Threads interface and offering students greater awareness of how usability and cognitive load shape their daily digital interactions.

II. METHODS

Research Design

This study employed a quantitative approach with a descriptive-correlational design. The descriptive component was used to depict the level of usability and cognitive load experienced by Threads users, while the correlational component tested whether a significant relationship exists between usability and cognitive load. This combination of designs was chosen because it allows the study both to characterize user perceptions and to statistically examine the extent to which usability quality relates to the cognitive load users report, providing developers with actionable insight into the interface's strengths and shortcomings.

Population, Sample, and Sampling Technique

The population comprised active, registered users of the Threads application in Indonesia in 2025, with a specific focus on students of Universitas Bhinneka PGRI who actively use Threads for social,

academic, or personal purposes. This group was selected because it represents a heterogeneous segment of users with varying levels of technological familiarity, and because students tend to be relatively open to adopting new social applications.

A sample of 200 active students who had used Threads for at least one month was selected using purposive sampling, a non-probability technique in which respondents are chosen according to criteria directly relevant to the research objective rather than to generalize across the entire user population. The inclusion criteria required respondents to be active students of Universitas Bhinneka PGRI, to have used Threads actively for at least one month, to have direct experience with the app's core features (posting, interacting with content, and managing account settings), and to be willing to complete the usability and cognitive-load questionnaire [18]. This approach yields more focused, experience-grounded data, though it necessarily limits the generalizability of the findings to populations with similar characteristics and may carry a degree of subjective bias inherent to self-reported perception data.

Research Variables

Two main variables were analyzed: usability (independent variable, X) and cognitive load (dependent variable, Y). Usability was operationalized through three sub-dimensions — effectiveness, efficiency, and user satisfaction — and measured with the ten-item SUS questionnaire, in which each item is rated on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) [1], [12]. Cognitive load was operationalized through the six NASA-TLX dimensions — mental demand, physical demand, temporal demand, performance, effort, and frustration level — each measured on a Likert scale that was subsequently transformed to a 0–100 range by multiplying raw scores by 20, consistent with the standard NASA-TLX measurement convention [2].

Hypothesis

Based on the theoretical framework, the study tested whether usability significantly influences cognitive load, hypothesizing that higher usability would be associated with lower cognitive load (i.e., an inverse relationship), reflecting the general expectation that a system that is easier to use should demand less mental effort from its users [7].

Data Collection and Analysis Technique

Data were collected through an online questionnaire distributed to the 200 selected respondents. Prior to hypothesis testing, the instruments were assessed for validity (Pearson product-moment correlation, item score against total score, compared against the r-table value for n = 200) and reliability (Cronbach's Alpha). Data were then examined for normality using the One-Sample Kolmogorov–Smirnov test on the regression residuals. Descriptive statistics summarized the level of usability and cognitive load, while Pearson correlation and simple linear regression analyses (processed in IBM SPSS) were used to examine the strength, direction, and predictive relationship between usability (X) and cognitive load (Y), including the coefficient of determination (R Square) and a t-test of the regression coefficient against the t-table at $\alpha = 0.05$.

III. RESULT AND DISCUSSION

Validity and Reliability Test

All ten SUS items and six NASA-TLX items produced calculated r-values greater than the r-table value of 0.138 (n = 200), indicating that every item was valid.

Table 1. Validity Test — SUS (Variable X)

Item	r calculated	r table	Status
Q1	0.178	0.138	Valid
Q2	0.690	0.138	Valid
Q3	0.221	0.138	Valid
Q4	0.778	0.138	Valid
Q5	0.226	0.138	Valid
Q6	0.695	0.138	Valid

Item	r calculated	r table	Status
Q7	0.173	0.138	Valid
Q8	0.701	0.138	Valid
Q9	0.225	0.138	Valid
Q10	0.715	0.138	Valid

Table 2. Validity Test — NASA-TLX (Variable Y)

Item	r calculated	r table	Status
MD	0.802	0.138	Valid
PD	0.797	0.138	Valid
TD	0.731	0.138	Valid
EF	0.803	0.138	Valid
P	0.175	0.138	Valid
RF	0.768	0.138	Valid

Reliability testing produced a Cronbach's Alpha of 0.728 for the 10-item SUS scale and 0.786 for the 6-item NASA-TLX scale. Both values exceed the commonly accepted 0.60–0.70 reliability threshold, confirming that both instruments were internally consistent and suitable for further analysis [13], [17].

Descriptive Statistics and Normality Test

Table 3. Descriptive Statistics

Variable	N	Minimum	Maximum	Mean	Std. Deviation
X (Usability)	200	27	46	33.01	4.287
Y (Cognitive Load)	200	8	28	17.94	4.716

The One-Sample Kolmogorov–Smirnov test on the unstandardized regression residual produced a test statistic of 0.046 and an asymptotic significance of 0.200, which is greater than 0.05. This indicates that the residuals are normally distributed, satisfying the normality assumption required for linear regression.

SUS Score Calculation

Responses from the 200 participants were scored using the standard SUS formula, producing individual scores on a 0–100 scale (multiplying the sum of adjusted item scores by 2.5).

Table 4. Sample of SUS Score Calculation (illustrative rows)

Respondent	Total Item Score	SUS Score
R01	29	72.5
R02	24	60.0
R04	28	70.0
R05	32	80.0
R200	33	82.5
Mean SUS Score	—	70.0875

The mean SUS score of 70.0875 places Threads in the "Good" category, above the commonly cited acceptability threshold of 68 [1], [12]. This indicates that, overall, users find the system's core functions reasonably operable without major obstruction, though the margin above the threshold is narrow enough to suggest meaningful room for improvement, particularly in navigation consistency and interface predictability.

NASA-TLX Score Calculation

Using the unweighted NASA-TLX model, the mean of the six workload dimensions (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) was computed for each respondent after transforming Likert responses to the 0–100 scale.

Table 5. Sample of NASA-TLX Score Calculation (illustrative rows)

Respondent	MD	PD	TD	P	EF	FR	Total	NASA-TLX Score
R01	60	40	60	80	60	40	340	56.67
R04	60	40	40	40	40	100	320	53.33
R05	80	100	100	20	80	100	480	80.00

Respondent	MD	PD	TD	P	EF	FR	Total	NASA-TLX Score
R200	40	20	40	80	20	40	240	40.00
Mean NASA-TLX Score								58.83

The mean NASA-TLX score of 58.83 falls into the "moderate" workload category. This suggests that although users can operate the platform's main features, they still require noticeable mental effort, attention, and adaptation to follow certain interaction flows — consistent with the pattern reported in comparable transportation and public-service application studies using NASA-TLX [2].

Regression and Hypothesis Testing

Table 6. Model Summary (Simple Linear Regression)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.474	0.225	0.221	4.163

Table 7. Regression Coefficients and Hypothesis Test

Model	B	Std. Error	Beta	t	Sig.
(Constant)	0.735	2.291	—	0.321	0.749
X (Usability)	0.521	0.069	0.474	7.574	0.000

The simple linear regression equation derived is **Y = 0.735 + 0.521X**. The regression coefficient of 0.521 indicates that each one-unit increase in usability is associated with a 0.521-unit increase in cognitive load. This coefficient is statistically significant, with a calculated t-value of 7.574 exceeding the t-table value of 1.972 at a significance level of 0.000 ($p < 0.05$), confirming that usability has a statistically significant effect on cognitive load. The R Square value of 0.225 indicates that usability explains 22.5% of the variance in cognitive load, while the remaining 77.5% is attributable to factors outside the scope of this study.

To verify the regression assumptions underlying this hypothesis test, the distribution and linearity of the residuals were examined visually.

Fig. 1 Histogram Dependent Variabel : Y

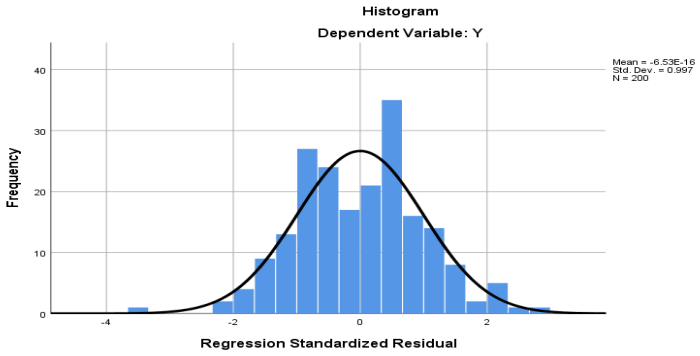

Fig. 1 below shows the histogram of the standardized residuals for the dependent variable (cognitive load), which approximates a normal, bell-shaped distribution centered near zero.

Figure 2 Grafik Normal P–P Plot Of Regression

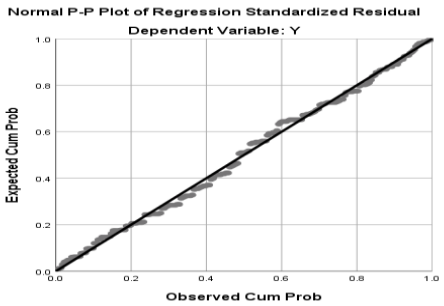

Fig. 2 shows the Normal P–P Plot of the regression standardized residuals. The residual points cluster closely around the diagonal reference line, following its direction, which further confirms that the residuals are normally distributed and that the normality assumption of the regression model is satisfied.

Pearson Correlation Test

Table 8. Pearson Correlation Between Usability (X) and Cognitive Load (Y)

	X	Y
X — Pearson Correlation	1	0.474**
X — Sig. (2-tailed)	—	0.000
Y — Pearson Correlation	0.474**	1
Y — Sig. (2-tailed)	0.000	—

Correlation is significant at the 0.01 level (2-tailed), N = 200.

The Pearson correlation coefficient of 0.474 ($p = 0.000 < 0.05$) indicates a statistically significant, moderate, positive relationship between usability and cognitive load — meaning that as usability scores increase, cognitive load scores also tend to increase.

IV. DISCUSSION

The usability measurement through SUS yielded a mean score of 70.0875, placing Threads within the "Good" interpretation category and above the generally accepted usability acceptability threshold of 68 [1], [12]. This confirms that the Threads interface is reasonably user-friendly and does not present major operational obstacles for its users, echoing prior evaluations of similarly text-oriented and mobile applications that report SUS scores in the "Good" range [4], [6], [9], [10].

The NASA-TLX result, by contrast, tells a related but distinct story. With a mean score of 58.83 in the "moderate" category, users are not entirely free of mental strain when interacting with Threads: recalling feature flows, processing a continuous stream of content, repeated scrolling, and adapting to intermittent notifications all contribute meaningfully to cognitive load [2], [7].

When these two results are considered together, they challenge the common assumption that a system rated favorably on usability will automatically relieve users of mental demand. This finding is particularly relevant to Threads' character as a text-based platform, where cognitive load arises not solely from interface complexity but also from the sheer volume and pace of content flow. The Pearson correlation test produced a coefficient of 0.474 ($p = 0.000$), confirming a meaningful, moderate-strength relationship between usability and cognitive load — and, notably, the relationship is positive rather than negative, meaning that increases in usability were accompanied by increases, not decreases, in cognitive load, a result that runs counter to the study's initial hypothesis of an inverse relationship [7].

The simple linear regression equation, $Y = 0.735 + 0.521X$, shows that each one-unit rise in usability corresponds to a 0.521-unit rise in cognitive load, confirmed by a t-value of 7.574 exceeding the t-table value of 1.972 at $p = 0.000$. The R Square of 0.225 indicates usability accounts for only 22.5% of the variance in cognitive load, leaving 77.5% explained by variables outside this study's scope. Taken together, these results affirm that usability does play a role in shaping cognitive load, but is far from the sole determinant — factors such as duration of use, social media habits, digital literacy, information density, and users' psychological state at the time of use likely contribute substantially as well [3], [7].

V. CONCLUSION

This study evaluated the usability and cognitive load of the Threads interface and examined the relationship and influence between the two constructs, based on data from 200 student respondents at Universitas Bhinneka PGRI Tulungagung. The usability level of Threads, measured through SUS, produced a mean score of 70.0875, indicating that users can generally understand available functions, navigate features without significant confusion, and complete their activities within reasonable limits — though gaps remain in visual consistency, navigation clarity, and operational efficiency. The cognitive load experienced by users, measured through NASA-TLX, produced a mean score of 58.83, a moderate workload driven primarily by mental demand, effort, and frustration, reflecting the continuous processing of fast-moving information, frequent notifications, and dynamic text-based conversations that Threads requires.

Statistically, usability and cognitive load are significantly and positively correlated ($r = 0.474$, $p = 0.000$), and usability significantly predicts cognitive load ($Y = 0.735 + 0.521X$, $R^2 = 0.225$), though usability alone explains only 22.5% of the variance observed. These findings suggest that while Threads is reasonably

usable, being easy to use does not automatically translate into low mental effort — a nuance developers and future researchers should account for when designing or evaluating similar text-based social platforms.

For application developers, continued refinement is recommended even though the usability score already exceeds the standard SUS threshold — particularly around interface consistency, simplified navigation, and smarter notification management to avoid unnecessarily amplifying cognitive load. For future researchers, given the modest R^2 value, expanding the scope of variables examined (e.g., frequency of use, digital literacy level, content density, and prior user experience) and considering complementary qualitative approaches would provide a more comprehensive understanding of the usability–cognitive load relationship in social media applications. For educational institutions, these findings may inform digital literacy programs, helping students better understand how digital interfaces work and how to manage cognitive load when using various digital platforms more effectively and efficiently.

REFERENCES

[1] Brooke, "SUS—A quick and dirty usability scale," in Usability Evaluation in Industry, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. London, U.K.: Taylor & Francis, 1996, pp. 189–194.

[2] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in Human Mental Workload, P. A. Hancock and N. Meshkati, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 139–183.

[3] J. Sweller, "Cognitive load during problem solving: Effects on learning," Cognitive Science, vol. 12, no. 2, pp. 257–285, 1988.

[4] Z. Mardinah, "Analisis usability testing website LMS PoltekSSN menggunakan metode System Usability Scale (SUS)," ***Indexa: Informatics and Computational Intelligent Journal***, 2024.

[5] Musyaffa, Priyanto, and A. Wijayanto, "Evaluasi user experience aplikasi KAI Access menggunakan SUS dan Cognitive Walkthrough," LEDGER Journal Informatics and IT, 2023.

[6] S. N. P. Dea, M. K. G. Suka, W. Aryani, and G. J. A. Indrawan, "Evaluasi usability aplikasi MyTelkomsel menggunakan metode System Usability Scale (SUS)," Journal of Computer Engineering, System and ***Science***, 2025.

[7] A. F. Putri and N. Suryani, "Pengaruh desain antarmuka terhadap beban kognitif pengguna aplikasi mobile," ***Jurnal Rekayasa Sistem dan Teknologi Informasi***, 2021.

[8] S. Hartini, M. Susilo, I. T. Handayani, H. Hafidzah, and U. Yuliani, "Analisis user experience pada aplikasi Threads menggunakan metode ***UEQ," Jurnal Ilmiah Teknik***, 2024.

[9] A. Akbar and R. Hanif, "Perancangan ulang UI & UX aplikasi Fello menggunakan System Usability Scale ***(SUS)," Jurnal Pendidikan dan Teknologi Indonesia***, 2025.

[10] G. Siahaan, R. Guntara, and D. Nuryadin, "Evaluasi usability website dashboard TMS Telkom Indonesia dengan metode usability testing dan SUS," ***Jurnal Minfo Polgan***, 2024.

[11] TechCrunch, "Threads now has more than 400 million monthly active users," Aug. 12, 2025. [Online]. Available: https://techcrunch.com/2025/08/12/threads-now-has-more-than-400-million-monthly-active-users/

[12] A. Bangor, P. Kortum, and J. Miller, "An empirical evaluation of the System Usability Scale," ***International Journal of Human-Computer Interaction***, vol. 24, no. 6, pp. 574–594, 2008.

[13] J. R. Lewis and J. Sauro, "The factor structure of the System Usability Scale," in Human Centered Design, Berlin, Germany: Springer, 2009, pp. 94–103.

[14] J. Nielsen, Usability Engineering. San Francisco, CA, USA: Morgan Kaufmann, 1993.

[15] International Organization for Standardization, Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts, ISO 9241-11:2018, Geneva, Switzerland, 2018.

[16] R. W. Proctor and T. Van Zandt, Human Factors in Simple and Complex Systems, 3rd ed. Boca Raton, FL, USA: CRC Press, 2018.

[17] S. C. Peres, T. Pham, and R. Phillips, "Validation of the System Usability Scale (SUS): SUS in the wild," in Proc. Human Factors and Ergonomics Society Annual Meeting, vol. 57, no. 1, 2013, pp. 192–196.

[18] U. Sekaran and R. Bougie, Research Methods for Business: A Skill-Building Approach, 7th ed. Chichester, U.K.: Wiley, 2016.