

Usability Evaluation of Grammar Learning Application Using SUS Method

Galuh Ihsan Nurkholis¹, Sri Mulyati²

Abstract

We proposed a mobile English grammar learning application based on the Formula 33 method to support structured, rule-based grammar instruction in a Mobile Assisted Language Learning environment. Many existing mobile language learning applications provide rich content but do not ensure that users can effectively follow systematic grammar learning workflows, which often leads to interaction difficulties and reduced learning focus. The proposed application guided learners through sequential stages of sentence identification, temporal analysis, and formula construction, and integrated quizzes along with an interactive learning assistant to support grammar exploration. This study utilized the System Usability Scale (SUS) to evaluate perceived usability, involving 32 participants who completed guided learning tasks before responding to the questionnaire. The evaluation produced a mean SUS score of 73.3, indicating a "Good" rating and an acceptable usability level above the standard benchmark. The results showed that users were able to understand and operate the structured learning process with minimal assistance, although first-time users experienced slight initial complexity. These findings demonstrate that an explicit, rule-based grammar learning approach can be effectively implemented in a mobile learning environment without introducing significant usability barriers.

Keywords:

System Usability Scale, Mobile-Assisted Language Learning, Formula 33, Usability Evaluation

This is an open-access article under the [CC BY-SA](#) license



1. Introduction

The rapid growth of mobile learning technologies transforms the way learners access educational content, including grammar instruction. Educational institutions increasingly adopt mobile platforms to support flexible and self-paced learning in both formal and open distance environments. However, many applications prioritize content delivery while overlooking usability aspects that directly influence learner engagement and effectiveness. Research on mobile learning technologies in open and distance learning environments shows that usability significantly determines user satisfaction, learning persistence, and system acceptance. When applications fail to provide intuitive navigation, clear instructions, and responsive interaction, learners experience cognitive barriers that reduce learning performance. Therefore, evaluating usability becomes essential to ensure that grammar learning applications truly support learners rather than hinder them [1], [3].

User experience (UX) research further emphasizes that mobile application success depends not only on functional completeness but also on perceived ease of use, efficiency, and satisfaction. A systematic literature review on mobile application UX identifies common usability challenges such as inconsistent interfaces, complex workflows, and inadequate feedback mechanisms. These issues become more critical in educational applications, where learners must concentrate on mastering complex concepts like grammar rules and

Corresponding Author: Sri Mulyati, Universitas Islam Indonesia (135230506@uii.ac.id)

¹ Sri Mulyati, Dept. Informatics, Faculty of Industry Technology, Universitas Islam Indonesia (135230506@uii.ac.id)

² Galuh Ihsan Nurkholis, Dept. Informatics, Faculty of Industry Technology, Universitas Islam Indonesia (21523128@students.uui.ac.id)

sentence structures. If the interface design imposes additional cognitive effort, learners may struggle to focus on instructional content. Thus, understanding user experience through systematic usability evaluation provides a strong foundation for improving grammar learning applications [2], [4].

In the context of language learning, mobile-assisted language learning (MALL) demonstrates strong potential to enhance accessibility and learner autonomy. Studies show that mobile environments support vocabulary acquisition, writing development, and interactive language practice. However, the effectiveness of educational technology applications depends heavily on how users interact with the system. A meta-analysis on educational technology for English learners reveals that technological tools improve learning outcomes only when they are designed with user-centered principles. Poor usability can reduce motivation and limit the pedagogical benefits of the application. Consequently, grammar learning applications must integrate instructional design with rigorous usability evaluation to maximize learning impact [8], [9].

Cognitive load theory also highlights the importance of interface simplicity in learning applications. When learners face complex layouts, unclear icons, or confusing instructions, they allocate cognitive resources to interface navigation rather than grammar comprehension. Empirical studies on mobile-assisted language learning environments demonstrate that well-designed systems reduce cognitive load and increase academic achievement and user acceptance. This finding implies that usability evaluation does not merely measure convenience but also influences learning efficiency and cognitive performance. Therefore, developers must assess whether grammar learning applications minimize extraneous cognitive load while supporting effective instructional interaction [15], [16].

To measure usability systematically, researchers widely adopt the System Usability Scale (SUS) as a reliable and efficient instrument. The SUS provides a quick yet robust method to evaluate perceived usability across various systems, including educational and mobile applications. Empirical validation studies confirm that SUS produces consistent results and correlates strongly with overall user satisfaction. Researchers also refine the interpretation of SUS scores by introducing adjective rating scales and examining its factor structure. These developments strengthen SUS as a standardized measurement tool for usability assessment in educational technology contexts [5], [6], [10], [12], [13].

Usability studies on mobile applications for phones and tablets further demonstrate that SUS effectively captures user perceptions across different device types. Comparative evaluations show that mobile-specific usability factors such as screen size, touch interaction, and portability. Since grammar learning applications commonly operate on smartphones, evaluating their usability requires attention to mobile interaction patterns. Research confirms that structured usability measurement helps developers identify weaknesses in navigation flow, response time, and task completion efficiency. Therefore, applying SUS to grammar learning applications aligns with established methodologies in mobile usability research [7], [11], [14].

Recent empirical studies evaluate educational applications, including language learning platforms, using SUS and related instruments. Evaluations of e-learning systems and language-based applications reveal that usability significantly influences learner satisfaction and continued usage intention. These studies show that even popular learning platforms may exhibit moderate SUS scores, indicating areas for interface and interaction improvement. Such findings highlight that educational value alone does not guarantee positive user experience. Instead, systematic usability evaluation becomes necessary to ensure that grammar learning applications meet user expectations in clarity, efficiency, and satisfaction [16], [17], [18].

Despite the growing body of research on mobile learning and SUS-based evaluation, limited studies focus specifically on grammar learning applications with detailed usability

analysis. Many investigations address general e-learning platforms or broader language learning systems without isolating grammar-focused interaction features. Grammar learning often involves structured exercises, rule explanations, quizzes, and feedback loops that require precise interface design. Without systematic evaluation, developers may overlook usability barriers that reduce learning effectiveness. Therefore, this study addresses the gap by conducting a structured usability evaluation of a grammar learning application using the System Usability Scale method, aiming to provide empirical evidence and practical recommendations for improving educational technology quality and learner-centered design [5], [6], [13].

2. Related Works

Previous studies examined the role of usability in mobile learning environments and confirmed its strong influence on learner engagement and performance. Rafiq et al. investigated the usability of mobile learning technologies in open and distance learning contexts and reported that ease of use and interface clarity significantly affected user satisfaction and continued usage intention [1]. Their study provided empirical evidence that usability problems reduced learning efficiency and increased user frustration. However, they focused on general mobile learning systems rather than specific language components such as grammar modules. As a result, their findings did not address how structured rule-based learning tasks interact with interface complexity in grammar-focused applications.

Research on user experience in mobile applications further strengthened the importance of systematic usability evaluation. Lu et al. conducted a systematic literature review and identified key UX dimensions, including effectiveness, efficiency, satisfaction, and emotional response [2]. They highlighted common weaknesses in mobile applications, such as inconsistent navigation and limited feedback mechanisms. Although their review offered a comprehensive conceptual framework, it did not provide empirical validation within specific educational domains. In particular, it did not analyze how usability metrics apply to grammar learning applications that require repetitive exercises and structured feedback.

Studies on mobile learning adoption in higher education also emphasized usability as a determinant of sustained engagement. Hameed and Sumari analyzed the adoption and continued usage of mobile learning platforms in unstable educational environments and found that perceived ease of use directly influenced acceptance and behavioral intention [3]. Similarly, Qazi et al. reviewed mobile learning during the COVID-19 pandemic and identified usability barriers as critical challenges affecting learning outcomes [4]. These studies demonstrated the strategic importance of usability but focused primarily on adoption and acceptance models. They did not conduct in-depth usability measurement using standardized instruments such as the System Usability Scale (SUS) for specific instructional applications.

In the field of language learning, Miangah explored mobile-assisted language learning and reported that mobile platforms enhanced learner autonomy and accessibility [8]. Xu et al. conducted a meta-analysis on educational technology applications for English learners and found positive effects on writing quality when digital tools supported structured learning activities [9]. Ozer and Kiliç examined mobile-assisted language learning environments and showed that well-designed systems reduced cognitive load and improved academic achievement [16]. These studies confirmed the pedagogical benefits of mobile technology. However, they primarily evaluated learning outcomes rather than systematically measuring perceived usability using validated usability scales.

The System Usability Scale (SUS) became one of the most widely applied instruments for usability evaluation. Brooke introduced SUS as a quick and reliable tool for assessing system usability across various domains [5]. Bangor et al. empirically validated SUS and

demonstrated its reliability and sensitivity in distinguishing acceptable and unacceptable systems [6]. They also proposed adjective rating interpretations to clarify the meaning of individual SUS scores [10]. Lewis and Sauro further analyzed the factor structure of SUS and confirmed its robustness for practical usability research [12], [13]. These foundational studies established SUS as a standardized method, yet they did not focus specifically on grammar-oriented educational applications.

Researchers later applied SUS to mobile applications to validate its effectiveness in mobile contexts. Kortum and Sorber measured the usability of mobile applications for phones and tablets and confirmed that SUS captured user perceptions across different device types [7]. Tullis and Stetson compared various usability questionnaires and concluded that SUS provided strong reliability with minimal administration effort [11]. Sauro also offered practical guidance on interpreting usability metrics and integrating quantitative evaluation into system development cycles [14]. Although these studies demonstrated methodological strength, they did not address subject-specific instructional systems such as grammar learning applications.

Recent empirical studies evaluated e-learning and language-based applications using SUS to measure user experience. Several researchers reported that even widely used educational applications achieved moderate SUS scores, indicating room for improvement in interface design and task flow. These studies highlighted strengths in content delivery but identified weaknesses in navigation clarity and feedback responsiveness. However, most evaluations focused on comprehensive e-learning systems rather than isolating grammar learning features, such as rule explanation modules, error correction feedback, and progressive difficulty levels. Consequently, they did not provide detailed insight into usability challenges unique to grammar instruction interfaces.

Therefore, prior research demonstrated the importance of usability in mobile learning, validated the reliability of the SUS instrument, and confirmed the educational benefits of mobile-assisted language learning. Nevertheless, existing studies rarely integrated these perspectives to evaluate a dedicated grammar learning application using a structured SUS-based approach. Most works either emphasized adoption factors, learning outcomes, or general system usability without focusing on grammar-specific interaction patterns. Therefore, this study addressed the identified gap by conducting a focused usability evaluation of a grammar learning application using the SUS method, aiming to provide empirical evidence and targeted design recommendations for improving grammar-focused educational technology.

3. Methodology

This study applies a quantitative usability evaluation approach to assess the suitability of a mobile grammar learning application based on the Formula 33 instructional model. In Mobile Assisted Language Learning (MALL), usability determines whether learners can effectively interact with instructional content and maintain engagement over time. Prior research shows that even pedagogically strong learning models fail when users perceive the interface as complex or unintuitive [6]. Therefore, this study treats usability evaluation as a central validation mechanism rather than a complementary analysis. The study selects the System Usability Scale (SUS) based on its relevance to the structured interaction design of the Formula 33 system, where users must identify sentence components, temporal references, and grammatical aspects before forming a final construction. In this context, the key issue extends beyond functional accuracy to whether users can cognitively perform the learning workflow without excessive mental effort, requiring evaluation in relation to users, tasks, and context of use [5].

The study measures perceived usability instead of performance-based metrics such as

task completion time or learning gain, since subjective measures enable standardized cross-system comparison [5]. SUS captures overall perceptions of complexity, confidence, and ease of use, and supports benchmarking across systems [10]. Comparative studies confirm that SUS produces reliable conclusions even with small samples and performs comparably to longer instruments [11]. Psychometric analyses further show that SUS measures both usability and learnability, a distinction crucial for grammar learning applications where users must quickly understand procedural steps without prior training [12]. Its high internal consistency and sensitivity to interface differences make SUS appropriate for evaluating the effectiveness of the Formula 33 mobile grammar learning environment.

1. Research Design

This study adopted a quantitative usability evaluation design using the System Usability Scale (SUS). Quantitative usability evaluation is widely applied in human-computer interaction (HCI) and educational technology research because it provides measurable and comparable indicators of user perception [13]. The SUS instrument was administered after the application reached functional stability, ensuring that all core features were operational during the evaluation process. The usability assessment was conducted following the iterative prototyping and black-box testing phases, which ensured that functional errors did not confound user perception during the usability test. This approach aligns with the best practices in usability engineering, where functionality validation precedes subjective usability assessment to obtain reliable user feedback. By evaluating the system at this stage, this study focused on the perceived ease of use, clarity of interaction, and overall user satisfaction rather than technical correctness alone.

2. Participants

A total of 32 respondents participated in the usability evaluation. This sample size is considered adequate for SUS-based studies, as prior research has shown that reliable usability insights can be obtained with samples ranging from 12 to 30 participants [14]. Participants were selected to represent the intended target users of the application, ensuring ecological validity of the evaluation results.

Table 1 Demographics of Participants

Attribute	Category	Frequency	Percentage (%)
Gender	Male/Female	n	%
Age Group	18-20 / 21-23 / >23	n	%
Educational Background	Undergraduate	n	%
Experience with Mobile Learning	Low / Moderate / High	n	%

The participant group consisted of undergraduate students with prior exposure to English grammar learning and general users with basic smartphone literacy. This composition reflects the application’s target audience, which includes learners seeking structured grammar support as well as non-specialist users interested in independent language learning. Including users with varying levels of grammatical knowledge is consistent with MALL research recommendations, as usability perception may differ based on prior domain familiarity [8].

3. Evaluation Procedure

The usability evaluation followed a structured task-based procedure designed to expose the participants to the core learning and interaction features of the application. Participants were first instructed to install the Formula 33 mobile application on their personal Android devices to ensure that the interaction occurred in a natural usage environment. They were then guided to navigate the learning formula module to familiarize themselves with the structure and presentation of the Formula 33 content.

Subsequently, participants were required to complete at least one quiz session to experience the rule-based grammar analysis workflow embedded in the application. To evaluate interactive learning support, participants were instructed to interact with the RAG-based chatbot by requesting grammar clarification related to the quiz or learning material. After completing these tasks, the participants completed the SUS questionnaire. This task-oriented evaluation approach is commonly recommended in usability studies because it enables users to form informed judgments based on direct system interaction. The SUS consists of ten items rated on a five-point Likert scale, alternating between positive and negative statements to minimize response bias and acquiescence effects[5].

4. SUS Score Method

The usability of the proposed mobile learning application was evaluated using the System Usability Scale (SUS), a standardized usability measurement instrument introduced by Brooke. The SUS questionnaire consists of ten items measured on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) and has been widely adopted in human-computer interaction research due to its reliability and efficiency in evaluating interactive systems [5].

To formally represent the SUS measurement, let Q_i denote the response score for item i , where each item has a discrete response value from 1 to 5.

$$Q_i \in \{1,2,3,4,5\}, \quad i = 1,2, \dots, 10$$

SUS includes alternating positively and negatively worded statements; therefore, response normalization is required before aggregation. Each item is converted into an adjusted usability contribution score s_i . Odd-numbered items are treated as positive statements, while even-numbered items are negative statements, following the original scoring procedure defined by Brooke [5]. The normalization ensures that higher values consistently represent better usability.

$$s_i = \begin{cases} Q_{i-1} & \text{if } i \text{ is odd} \\ 5 - Q_i & \text{if } i \text{ is even} \end{cases}$$

After normalization, the adjusted scores from all ten items are aggregated to obtain the raw usability score.

$$S_{total} = \sum_{i=1}^{10} S_i \quad (1)$$

Since the maximum possible raw score is 40, the total score is multiplied by a constant factor of 2.5 to transform it into a usability score within a 0–100 scale. This scaling allows easier comparison across systems and is recommended in SUS statistical interpretation guidelines[13].

$$SUS = 2.5 \times S_{total} \quad (2)$$

Thus, the final SUS score is defined in the interval:

$$0 \leq SUS \leq 100$$

To obtain an overall usability evaluation, the SUS score of each respondent was averaged across all participants.

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n SUS_j \quad (3)$$

Where \bar{x} is the mean SUS score, SUS_j is the SUS score of respondents j , and n is the number of respondents. For clarity and reproducibility, the mathematical symbols used in the SUS measurement model are summarized in Table 1.

Table 1. The Mathematical Notation of SUS Equation

Notation	Description
Q_i	Response score of the i -th SUS questionnaire item (1-5 Likert score)
i	Item index (1-10)
s_i	Normalized usability contribution score of items i
S_{total}	Total normalized score obtained from all SUS items
SUS	Final System Usability Scale score
SUS_j	SUS score of respondent j
j	Respondent index
n	Number of respondents
\bar{x}	Mean SUS score from all respondents

The interpretation of SUS scores is based on established benchmarking guidelines that map numerical usability scores into qualitative usability categories. The scale includes grade ratings, adjective ratings, and acceptability levels, enabling a more comprehensive understanding of usability performance beyond numerical values. As illustrated in Fig. 1, higher SUS scores correspond to better perceived usability, ranging from unacceptable systems to highly usable systems.

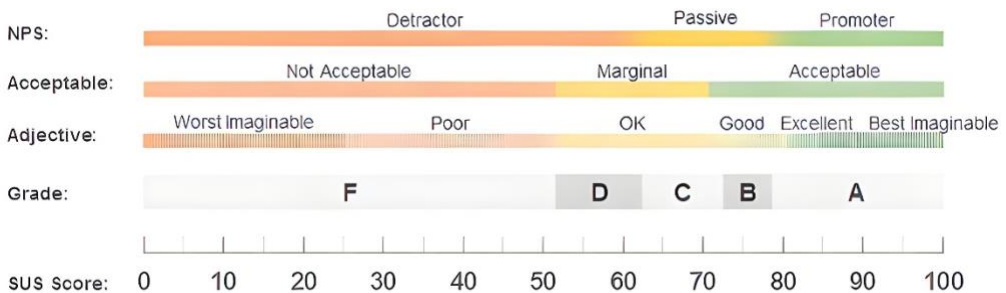


Fig. 1 SUS Score Interpretation

4. Result

This section presents the results of the usability evaluation using the System Usability Scale (SUS). The findings are reported in two parts: the overall SUS score, which reflects users' general perception of the application's usability, and an item-level analysis that provides insight into specific interaction aspects influencing user experience.

1. SUS Score

The usability evaluation yielded a mean SUS score of 73.3, indicating that users were generally able to interact with the application without significant operational difficulty [13]. More importantly, the result suggests that the sequential reasoning process required by the Formula 33 method can be performed through a mobile interface. Unlike drill-based language applications, the learning procedure requires users to identify sentence type, temporal reference, and grammatical aspect before constructing the final formula. The obtained usability score therefore indicates that the interface supports the cognitive workflow of the grammar method rather than obstructing it. In SUS benchmarking literature, scores above the baseline value of 68 are considered above average and indicative of satisfactory usability performance[6].

Table 2 SUS Test Result Score

Metric	Result
Number of respondents	32
Mean SUS score	73.3
Minimum score	45
Maximum score	97.5
Grade scale	B
Adjective rating	Good
Acceptability range	Acceptable

The System Usability Scale (SUS) evaluation involved 32 respondents. The results yielded a mean SUS score of 73.3, which corresponds to a Grade B rating, an adjective rating of "Good," and falls within the Acceptable usability range[6]. Further analysis of the SUS score distribution reveals that the dominant usability category is "Excellent" (SUS > 80), representing 37.5% of the respondents. This indicates that a substantial proportion of users perceived the application as highly usable. Meanwhile, 21.9% of respondents rated the system within the "Good" category, while 31.3% fell into the "Marginal" range and only 9.3% perceived the usability as "Poor." Overall, the dominance of the "Excellent" category supports the overall mean SUS score and suggests that the application demonstrates strong usability performance, with minor usability issues perceived by a smaller subset of users.

To better interpret the usability result, the obtained SUS score was compared with established SUS benchmark values. Previous studies report that the average SUS score across a wide range of interactive systems is approximately 68, which represents the baseline usability level. The score of 73.3 obtained in this study exceeds this baseline, indicating that the application performs above the average usability level of general software systems. According to SUS interpretation guidelines, scores between 70 and 80 fall within the "Good" usability region and are associated with systems that users can operate with minimal assistance and acceptable learning effort. Therefore, the result demonstrates that the mobile implementation does not hinder the structured

learning procedure required by the Formula 33 instructional method.

The result is particularly significant because the evaluated application requires users to follow a sequential analytical learning process rather than simple interaction tasks. Grammar analysis involves identifying sentence components and temporal aspects before constructing the final formulation, which increases cognitive demand during early interaction[15], [16]. Despite this requirement, users were still able to achieve above-average usability perception. This indicates that the interface successfully supports the learning workflow instead of introducing operational barriers. Consequently, the usability result serves as validation that the instructional method can be practically applied in a mobile learning environment rather than merely demonstrating interface convenience.

Table 3 Distribution of SUS Scores by Category

SUS Category	Score Range	Frequency (n)	Percentage (%)
Excellent	> 80	12	37.5%
Good	68 – 80	7	21.9%
Marginal	50 – 67	10	31.3%
Poor	< 50	3	9.3%
Total		32	100%

The distribution of scores shows that although a substantial proportion of users experienced high usability, a notable group reported marginal usability. This variation indicates that the difficulty perceived by some participants may be related to understanding the analytical learning steps rather than navigating the interface. The Formula 33 method requires users to perform structured grammatical reasoning, and first-time users may initially attribute cognitive difficulty to the application interface. Therefore, the presence of marginal scores does not necessarily indicate interface failure but reflects the learning curve associated with adopting a rule-based grammar learning strategy in a mobile environment.

Overall, the findings indicate that the application enables users to carry out grammar analysis tasks independently. The acceptable usability level demonstrates that the mobile implementation does not interfere with the structured reasoning process required by the instructional method. Consequently, the usability result serves as evidence that the Formula 33 learning model can be operationalized in a mobile learning environment, supporting self-directed grammar learning rather than merely confirming interface convenience.

2. Item-Level Analysis

Further analysis of individual SUS items provides a more detailed understanding of user interaction with the application. High mean scores were observed on items related to ease of learning, confidence in system use, and perceived integration of features. These results indicate that users were able to quickly understand how to operate the application and felt comfortable navigating between learning modules, quizzes, and chatbot-based assistance. Prior studies note that strong performance on these SUS items is often associated with systems that offer coherent learning flow and consistent interaction patterns [13]. Slightly lower scores were recorded on items associated with perceived system complexity, particularly those reflecting initial impressions during first-time use. This suggests that although the application is generally usable, some users may require brief familiarization before fully understanding the interaction logic. Such findings are common in rule-based educational systems, where structured workflows may initially appear complex despite

being logically consistent [6]. Overall, the item-level results indicate that the learning flow is clear and well-structured, while minor interface simplification or introductory guidance could further enhance first-time user experience.

Table 4 Item-Level SUS Statistics

Item	Mean	Std. Dev.	Interpretation
Q ₁	3,87	0.69	Moderate
Q ₂	2,18	1.01	Low
Q ₃	4,12	0.99	High
Q ₄	2,36	1.27	Low
Q ₅	4,18	1.01	High
Q ₆	2	0.86	Low
Q ₇	4,03	0.98	High
Q ₈	1,84	0.83	Low
Q ₉	4,06	1.05	High
Q ₁₀	2,39	1.36	Low

5. Discussion

The System Usability Scale evaluation shows that the Formula 33 mobile grammar application achieves an average score of 73.3, which falls within the “Good” and Acceptable range, indicating that users can effectively operate the system without extensive prior training. The integration of a Retrieval-Augmented Generation (RAG) chatbot enhances perceived usability by providing contextual grammar clarification and increasing user confidence through consistent, document-based responses. Although slightly lower scores on perceived complexity suggest an initial learning curve, this result aligns with Cognitive Load Theory, as first-time users must adapt to a structured, rule-based grammar analysis process that requires simultaneous processing of sentence structure and tense [15], [16]. Importantly, the findings validate not only the interface design but also the practical implementation of the Formula 33 instructional model in a Mobile Assisted Language Learning environment. Since SUS measures global usability, including learnability, confidence, and complexity [5], [12], [13], the acceptable score confirms that the application supports analytical grammar learning without hindering user engagement, thereby demonstrating the feasibility of applying a structured rule-based approach in mobile learning contexts [4], [8].

From a practical perspective, the findings indicate that the Formula 33 mobile application is suitable for both formal and informal learning contexts, as its “Good” usability rating suggests that learners can operate the system with minimal guidance in classrooms, self-directed study, or supplementary learning environments. In Mobile Assisted Language Learning (MALL), such usability ensures that learners focus on grammar content rather than interface difficulties, thereby promoting sustained engagement and autonomous learning [8]. The results also confirm that the System Usability Scale (SUS) serves as an effective instrument for iterative improvement because it provides standardized and interpretable scores that developers can use across development cycles to monitor usability trends and evaluate design refinements [13], [14]. From a design standpoint, the findings highlight the need to strengthen onboarding and tutorial features to ease first-time user adaptation, since early interaction experiences significantly influence long-term acceptance in mobile learning systems [6].

This study has several limitations that should be considered when interpreting the findings. First, the usability evaluation involved a moderate sample size, which, although

sufficient for System Usability Scale (SUS) analysis, may limit the generalizability of the results to broader learner populations. Previous usability research indicates that SUS can yield reliable insights with relatively small samples; however, larger and more diverse participant groups could provide a more comprehensive understanding of usability across different user profiles.

Second, the evaluation focused on short-term interaction with the application and did not incorporate longitudinal measurement of learning outcomes or sustained usage behavior. While SUS effectively captures users' immediate perceptions of usability, it does not directly assess how usability influences long-term engagement, learning retention, or performance improvement. As noted in prior studies, prolonged interaction and repeated use may reveal usability challenges that are not apparent during initial evaluation sessions.

6. Conclusion

This study evaluated the usability of the Formula 33 mobile grammar learning application using the System Usability Scale (SUS) with 32 participants who interacted with learning modules, quizzes, and a chatbot-assisted inquiry feature. The evaluation produced a mean SUS score of 73.3, which corresponds to Grade B, an adjective rating of "Good," and an acceptable usability level. This score exceeds the standard SUS benchmark of 68, indicating that the application performs above the average usability level of general interactive systems.

The findings show that users understood and followed the structured grammar learning procedure with relatively low interaction difficulty, despite the requirement for sequential analytical reasoning. Item-level analysis reveals strong learnability and user confidence, while slightly lower scores relate primarily to initial familiarization during first-time use. These results confirm that a rule-based and structured instructional method can be implemented effectively in a mobile learning environment without creating significant usability barriers.

Therefore, the study demonstrates that usability evaluation not only assesses interface quality but also validates the practical applicability of the Formula 33 instructional approach in Mobile-Assisted Language Learning. The findings emphasize the importance of measuring usability when adapting pedagogical models into digital platforms. Future research should extend beyond usability assessment to examine learning effectiveness through pre-test and post-test experimental designs, integrate adaptive and personalized feedback features, and compare the Formula 33 approach with other mobile language learning platforms to evaluate long-term instructional impact.

References

- [1] N. Rafifing, J. Mosinki, A. Mabina, B. E. Othomile, and O. Mphole, "Usability of mobile learning technologies in open and distance learning," *Journal of Information Systems and Informatics*, vol. 7, no. 1, pp. 138–157, Mar. 2025, doi: 10.51519/journalisi.v7i1.989.
- [2] G. Lu, S. Qu, and Y. Chen, "Understanding user experience for mobile applications: A systematic literature review," *SN Applied Sciences*, Jun. 2025.
- [3] R. A. G. Hameed and P. Bin Sumari, "Adoption and continued usage of mobile learning virtual platforms in Iraqi higher education in an unstable environment," *International Journal of Information Management Data Insights*, vol. 4, no. 2, Nov. 2024.
- [4] Qazi, J. Qazi, K. Naseer, N. Hasan, G. Hardaker, and D. Bao, "M-learning in education during COVID-19: A systematic review of sentiment, challenges, and opportunities," *Heliyon*, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32638.
- [5] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*. London, U.K.: Taylor & Francis, 1996, doi: 10.1201/9781498710411-35.

- [6] Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the System Usability Scale," *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, Aug. 2008, doi: 10.1080/10447310802205776.
- [7] Bangor, P. T. Kortum, and J. T. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.
- [8] P. Kortum and M. Sorber, "Measuring the usability of mobile applications for phones and tablets," *International Journal of Human-Computer Interaction*, vol. 31, no. 8, pp. 518–529, Aug. 2015, doi: 10.1080/10447318.2015.1064658.
- [9] J. R. Lewis and J. Sauro, "The factor structure of the System Usability Scale," in *Human Centered Design*, LNCS 5619, 2009, pp. 94–103.
- [10] J. R. Lewis, "The System Usability Scale: Past, present, and future," *International Journal of Human-Computer Interaction*, vol. 34, no. 7, pp. 577–590, Jul. 2018.
- [11] T. S. Tullis and J. N. Stetson, "A comparison of questionnaires for assessing website usability," in *Proc. Usability Professionals Association Conf.*, 2004.
- [12] J. Sauro, *A Practical Guide to Measuring Usability: 72 Answers to the Most Common Questions About Quantifying the Usability of Websites and Software*. Denver, CO, USA: Measuring Usability LLC, 2010.
- [13] T. M. Miangah, "Mobile-assisted language learning," *International Journal of Distributed and Parallel Systems*, vol. 3, no. 1, pp. 309–319, Jan. 2012, doi: 10.5121/ijdps.2012.3126.
- [14] Z. Xu, M. Banerjee, G. Ramirez, G. Zhu, and K. Wijekumar, "The effectiveness of educational technology applications on adult English language learners' writing quality: A meta-analysis," *Computer Assisted Language Learning*, vol. 32, no. 5–6, pp. 1–26, 2019, doi: 10.1080/09588221.2018.1501069.
- [15] Ozer and F. Kiliç, "The effect of mobile-assisted language learning environment on EFL students' academic achievement, cognitive load and acceptance of mobile learning tools," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 14, no. 7, pp. 2915–2928, 2018, doi: 10.29333/ejmste/90992.
- [16] M. Kahfi M. S. and E. A. M. Sampetoding, "Evaluation of Duolingo user experience using the System Usability Scale (SUS) and End User Computing Satisfaction (EUCS) methods," *Jurnal Media Infotama*, vol. 21, no. 1, pp. 194–201, Apr. 2025, doi: 10.37676/jmi.v21i1.7812.
- [17] N. Rahayu, "UX analysis on e-learning application using the System Usability Scale (SUS) method," *Jurnal Komputer*, vol. 3, no. 2, pp. 49–54, 2025, doi: 10.70963/jk.v3i2.105.
- [18] M. R. Pratama, J. Umam, and R. Yakok, "Usability testing of the iJateng application using the System Usability Scale method," *Jurnal Sistem Informasi, Manajemen dan Teknologi Informasi*, vol. 2, no. 1, 2023, doi: 10.33020/jsimtek.v2i1.556.
- [19] Y. U. Budiman, "Usability evaluation of the Linkterasi application using the System Usability Scale (SUS) method," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, 2025, doi: 10.31004/riggs.v4i3.2812.
- [20] Kholin and J. T. K. Perangin Angin, "Usability and UX of Ruangguru mobile: A 120-user evaluation with SUS and UEQ," *Sinkron: Jurnal Penelitian Teknik Informatika*, vol. 9, no. 4, 2024.