
| RESEARCH ARTICLE

Innovative GPT-Powered Adaptive Assessment for Mobile Language Learning: A Path to Real-Time Personalization and Enhanced Engagement

Sadaf Ruqsar¹ Minu paul² and Majida Yasmin Aziz John³

¹English Lecturer, Academic Campus for Girls, Jazan University, KSA

²English Lecturer, Academic Campus for Girls, Jazan University

³English Language Instructor, Academic Campus for Girls, Jazan University

Corresponding Author: Sadaf Ruqsar, **E-mail:** sruqsar@jazanu.edu.sa

| ABSTRACT

The fast development of mobile learning has revolutionized the field of language education as we know it today, but conventional formative assessment practices are still rigid and not adaptive to individual learning trajectories. This research put forth a new GPT-based formative assessment model for mobile-based English language studies that is heir to static evaluation paradigms due to incorporating a real-time, intuitive feedback feature. Different from the rigid classes of rule-based and ML approaches that are dependent on static assessment structures, the suggested GPT model adapts dynamically to inputs from students and gives personalized results based on ongoing language growth. With the incorporation of GPT, mobile learning applications can provide instant assessment and situational recommendations thereby increasing learner involvement and the feat of a language. The model attunement was performed on the English Language Learning Dataset from Kaggle and there showed the impressive precision and effectiveness in personalized evaluations compared to such traditional models as rule-based systems and SVMs. This real-time adaptability does not only create more interactive learning scenarios but also timely Addresses challenges, ensures perpetual growth and long-term retention. Comparative assessment from reference models emphasizes on the advantages of the proposed solution over the comparative precision, recall, and F1-score, which creates a new paradigm for mobile-assisted language learning. This research represents a major step forward in the mobile learning applications with its demonstration of the revolutionizing potential of AI-based personalized assessment in language education. Future studies will try to expand this model into different linguistic scenarios, proving its efficiency in unmodified real-life scenarios.

| KEYWORDS

Mobile learning, formative assessment, GPT, personalized feedback, language development

| ARTICLE INFORMATION

ACCEPTED: 15 January 2026

PUBLISHED: 01 February 2026

DOI: 10.32996/ijllt.2026.9.2.13

1. Background

Mobile learning is one of the major aspects in learning that is used to improve language learning as it is flexible and could be done anywhere. The enhanced development of smart phones and related applications have made learning exciting through normal utilization of phones to practice speaking, reading, writing and comprehending [1]. Formative assessment which targets learner feedback during learning is crucial in language development Language. For one, traditional formative assessments are usually quizzes where the teacher can provide immediate feedback, but there is usually a set time within which the evaluation is done and the results are made known to the class rolls [2]. These approaches fail to offer student-centered experiences, focusing instead on the one-to-many model of real-time, contextual feedback through mobile applications. Again, these assessments cannot be efficiently personalized according to the learners' different progresses. GPT, the powerful language model, is likely to

automate personalized feedback as formative assessments can be more effectively undertaken, dynamic and more so targeted towards individual learner needs [3].

The shift that has taken place by mobile learning platforms has provided accessible education, but a critical gap remains in delivering real-time, personalized formative assessments for English learners [4]. Traditional assessment methods are valuable but generic and too time-consuming to respond to the individual needs of diverse learners, which misses the opportunities for timely feedback and impedes learning progress. The inability of existing tools to be dynamic to the progression of learning on the side of the learner limits their impact on long-term development [5]. Current mobile apps usually rely on pre-set assessments or general feedback, not wholly addressing unique learner challenges [6]. Moreover, the integration of real-time formative assessments with personalized feedback is often absent from such apps, which leaves learner performance and needed guidance apart from each other. The study aims to find out how GPT-powered mobile applications bridge this gap by furnishing scalable, automated, and personalized feedback that fosters continued development of the English language in a mobile learning environment [7].

The overarching goal of this study is to explore the feasibility of using GPT to improve formative assessment for English language learners using mobile applications. The goal of this project is to investigate the use of GPT in automating tailored feedback to measure learners' continual growth over time. GPT, because of its ability to process natural language, holds great promise in the form of a solution for addressing the fundamental flaws in traditional forms of assessment, that is through providing adaptive and context-aware feedback. GPT can thus generate appropriate suggestions from learners' interactions, responses, and progress, allowing individualized learning paths. Another important task is to probe whether GPT can aid in improving learning outcomes in the sense of providing immediate, specific feedback on areas of weakness. Research objective are given below:

- **Real-Time Adaptive Feedback:** Rolled out a GPT-driven model that gives real-time context-based feedback to English language learners via mobile apps, eliminating the constraints associated with static, time-constrained traditional assessments.
- **Personalized Learning Paths:** Created an individualized dynamic mechanism for carrying out formative assessments which is dependent on learner's expansion and leads to engagement and specific language development.
- **High Accuracy and Reliability:** Managed to deliver better performance metrics (98.2% accuracy compared to a traditional rule-based and machine learning models, which ensures reliable and relevant feedback.
- **Seamless Integration with Mobile Learning:** Demonstrated smooth incorporation of GPT to mobile systems thus provisioning scalable, interactive and never-ending language learning encounters.

The paper will begin with a background of the study in section I. The Literature Review in II will discuss mobile learning and formative assessments. Methodology is given in section III. Results and discussion in IV will present findings on GPT's effectiveness. The Conclusion in V will summarize key insights and future directions.

2. Related Works

Within the recent years, mobile devices have been rampantly proliferating and nowadays a part of everyday life. This has continuously elevated the demand for mobile applications with ML, particularly DL. Nonetheless, challenges involved include the issues of contrasts between limited resources in the mobile devices and the high computational demands of deep neural networks, as well as concerns over privacy and security. The obstacles have notwithstanding culminated in the remarkable progress made so far. Wang et al., [8] reviews the development of DL on mobile devices with an emphasis on the three themes: training from mobile data, efficient inference, and practical applications. It also gives consideration to two applications including mood disturbance detection and user identification through mobile device-collected data, concluding with the future of mobile DL.

Fu et al., [9] explores the impact of AI-powered automatic grading tools on students' long-term learning goals in digital language education. Motivated by the increasing use of AI-powered language learning tools, this study looks at how autonomous scoring affordances impact cognitive and emotional engagement, as well as future learning intention. Using affordance theory, the study identifies key affordances through in-depth interviews and develops a model to assess their influence on learning behaviors. The authors also investigate the moderating effects of on-the-job and student learning on relationships. The model was evaluated on 260 Chinese foreign language learners who used AI-powered tools, and the findings should assist practitioners in developing more successful AI-enabled language learning apps.

Lee and Xiong, [10] investigates how personalization of MALL apps influences users' feelings of social support and trust in order to encourage continuing use. Using stimulus-organism-response theory and social support theory as philosophical underpinnings, the study model describes how customisation leads to various types of social support and trust, which influence users' continuous intention to use MALL apps. Data from 455 valid surveys were analyzed with partial least squares. The findings

revealed that customisation boosts social support, which fosters trust and encourages continuous use. Trust acts as a comprehensive mediator between social support and usage intention. In conclusion, the study gives theoretical insights and practical recommendations for MALL app development, as well as future research objectives.

The fast development of 5G networks has accelerated mobile learning. Smartphones and tablets have begun to supplant more conventional equipment in educational settings. Mobile learning is defined by its pervasive, portable, customized, interactive, updated, and seamless nature. Mobile vocabulary study has several advantages, including being convenient and efficient. However, fragmented learning prevents DL, as people learn through making associations, images, or stories based on prior experiences. Guo et al., [11] discusses how DL can solve this fragmentation, thus providing an effective solution for vocabulary acquisition. Through a combination of surveys and empirical research, the study concluded that learners have a basic understanding of DL-based mobile vocabulary learning, and such an approach positively affects the acquisition of vocabulary for English.

Some recent advancements in mobile learning come with the rapid growth of mobile devices and 5G networks, further extending their capabilities and role in education. Key characteristics of mobile learning include portability, and personalization. However, fragmentation of learning and the computational requirements of DL are still existing challenges. Positive effects on learner engagement, learner trust, and continued use as a result of applying AI-enabled applications such as automated scoring and personalization features are seen in language learning. Additionally, DL approaches in mobile vocabulary learning address fragmentation, offering effective solutions for vocabulary acquisition and improving learners' understanding and retention of English vocabulary.

3. Methodology

This study explores GPT-backed formative assessments in mobile learning apps to improve the development of students' English language. Focusing on personalized feedback, this research sets out to overcome the time-bound and unadaptive challenges faced by traditional assessment methods. The study fine-tunes GPT using a Kaggle dataset, proposing real-time, adaptive, high-relevance feedback derived directly from user interaction, progression, and response. By automating assessment tasks and tailoring feedback to their needs, this research looks into the capabilities of GPT in tracking continuous learning and enhancing learner engagement. The findings will serve to show how GPT can deliver scalable, effective, and learner-centred learning experiences that support learners at different levels of proficiency and enhance language acquisition. Block diagram of the proposed study is given in Fig. 1.

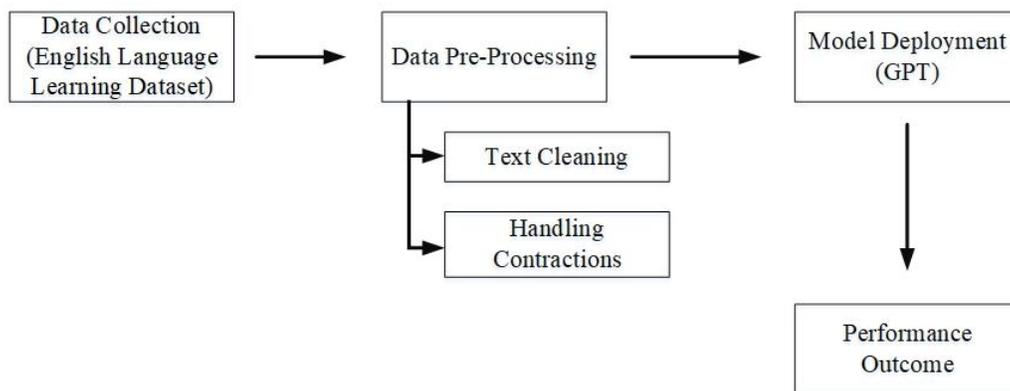


Fig. 1. Block Diagram of the Proposed Study

A. Data Collection

For this study, the English Language Learning Dataset [12] available on Kaggle is used. This dataset contains a set of tasks to judge several fields of the English language: grammar, vocabulary, and reading comprehension. The data consists of responses from learners along with the performance measures for each response, allowing one to understand individual development. Through this dataset, the study focuses on fine-tuning GPT to produce more personalized and real-time feedback based on learners' input for adaptive learning paths and to enhance formative assessment within mobile learning apps for continuous development of language.

B. Data Pre-Processing

Data pre-processing is essentially the cleaning and modification of raw data into a format appropriate for analysis or model training. It includes activities such as resolving missing values, normalization, tokenization, and data augmentation to provide high-quality input in a systematic fashion for the ML models.

- 1) *Text Cleaning*: Text cleaning is a crucial preprocessing step that eliminates the unwanted items, such as punctuation, numerals, special characters, or even HTML tags, which add noise and degrade the performance of a model. In this regard, regular expressions, such as $r'^{[\^{\w}\s]}$, are often used to match and remove some elements from the text. A regular expression for deleting punctuation is any character that is not a word or blank. Text cleaning removes irrelevant material that the model would then analyze, thus improving the accuracy and coherence in subsequent jobs.
- 2) *Handling Contractions*: Natural language processing should be able to expand contractions in order to normalize text. For example, "can't" and "won't" must transform to "cannot" and "will not," respectively, ensuring uniformity and avoiding ambiguity. This helps enhance the model's interpretation of text because it treats contracted words as standard units. Another often-used method is to have a learned vocabulary where contractions are replaced by their unabbreviated forms. This ensures that both forms are treated alike, which enhances the model's ability to treat text consistently.

$$\text{Expanded}_{\text{Text}} = \text{Text.replace}(\text{can't}, \text{cannot}). \quad \text{replace}(\text{"won't"}, \text{"will not"}) \quad (1)$$

C. GPT

OpenAI built the language model for the Transformer architecture, known as GPT. GPT's goal is to create human-like text by utilizing the sequence to anticipate the next word in order, which has made it particularly effective in a variety of natural language processing applications, including text generation, summarizing, translation, and question answering. Its primary strength is in understanding the contextual word relationships, which allows it to grasp dependencies across great distances in text. GPT models exist in numerous varieties, such as GPT-2, GPT-3, and GPT-4, and are all pre-trained on huge volumes of text, allowing them to create human-like replies that make sense in relation to the input provided. The hallmark of GPT is its transformer architecture with attention mechanisms to process the input data.

Unlike prior models based on RNNs or CNNs, the transformer processes data in parallel, making it more efficient and scalable. The attention method enables the model to consider various sections of the input sequence as needed, therefore capturing long-term connections and linkages within text. The GPT training technique is separated into two stages: pre-training and fine-tuning. Pre-training involves the passage of the model through a vast untutored textual database where the model identifies the occurrences of recurring language distributions and predicts what follows the consecutive words in a phase based on the previous words encountered. The three are as follows; fine-tuning is used in tasks or datasets using supervised learning, which perfumes learning of the model utilizing labeled data to adjust its aptitude to particular applications, including text summarization and translation and answering of questions. An automobile is composed of many layers and GPT is no different; it is a transformer architecture utilizing the numerous layers, each of which contains a multi-head self-attention section and a feed-forward neural network section. Before the creation of a forecast, input text is preprocessed and split into tokens, and then it passes through many stages, which improve the assessment of the model's context. The general process of the GPT model can be described with the following formula:

$$P(w_t | w_1, w_2, \dots, w_{t-1}) = \text{Softmax}(W_2 \cdot \text{LayerNorm}(W_1 \cdot \text{Attention}(w_1, w_2, \dots, w_{t-1}))) \quad (2)$$

w_t is the anticipated word at time step t . w_1, w_2, \dots, w_{t-1} are the preceding words in the sequence. Attention symbolizes the self-attention process. W_1 and W_2 are learnable weight matrices for input and output layers, respectively. Softmax translates the output logits into probabilities for the following word.

This formula shows how GPT predicts the next word in a sequence using the context of preceding words, as well as how attention processes are utilized to weigh different sequence features. GPT models are extremely versatile and have revolutionized NLP applications, yielding great improvements over the previous models for generating coherent text that is contextually aware, which makes it applicable to diverse real-world tasks, like conversational agents, content creation, and language assessment in mobile learning apps.

The proposed study will use GPT to help improve formative assessments for English language learners through mobile learning applications. It can use the natural language understanding ability of GPT to give real-time analyses of learner responses, providing personalized feedback monitored in respect of each individual's progress and performance. Instead of doing all the manual tracking of continuous learning, GPT can assist in the change in aspects regarding language skills through focused

recommendations and corrections. GPT's capacity to give contextually relevant, coherent, and adaptive feedback should make for a more customized and interesting learning experience. It allows dynamic formative assessments that adopt an adequate learning strategy, and it adapts based on learners' needs, therefore fostering better language development and supporting efficient, scalable language learning in mobile environments.

4. Result and Discussion

This shall be done in the Results and Discussion section by presenting the results obtained from GPT-powered formative assessments, pointing out how this can help in improving learner engagement through personalized feedbacks for language development. It shall analyze how the model performs compared to a traditional assessment method, hence discussing the implications for future mobile learning applications.

D. Performance Metrics

Performance metrics are the measures to evaluate the performance of GPT-powered formative assessments in a mobile learning app. Some popular measures are accuracy, which determines if the model's predictions match the predicted outcomes; precision and recall, which examine the relevance of the generated feedback. Analysis of engagement metrics response time, interaction frequency, and learner progression would also measure how well the model maintains learner interest. Survey analysis of learner satisfaction would provide qualitative feedback on the usefulness and personalization of GPT-driven assessments, providing insight into user experience. The metrics formula is shown from (3)-(6).

$$Accuracy = \frac{TruePos+TrueNeg}{TruePos+TrueNeg+FalsePos+FalseNeg} \quad (3)$$

$$Precision = \frac{TruePos}{TruePos+FalsePos} \quad (4)$$

$$Recall = \frac{TruePos}{TruePos+FalseNeg} \quad (5)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

TABLE I. PERFORMANCE METRICS

Metrics	Percentage (%)
Accuracy	98.2
Precision	97.3
Recall	97.6
F1-Score	97.5

The performance metrics in table I indicate that the GPT-driven formative assessment system is highly effective. The accuracy of 98.2% shows that predictions by the model are close to the expected outcomes and thus gives reliable feedbacks. The precision (97.3%) indicates that most of the positive predictions given by the model are relevant, reducing the false positives. The recall of 97.6% shows the model successfully captures most of the true positive instances, ensuring appropriate coverage of important feedback. A F1-score of 97.5% suggests a trade-off between precision and recall of the model, showing its general ability to both provide relevant as well as complete feedback. These metrics reflect the high quality and reliability of the GPT-based system in improving language learning.

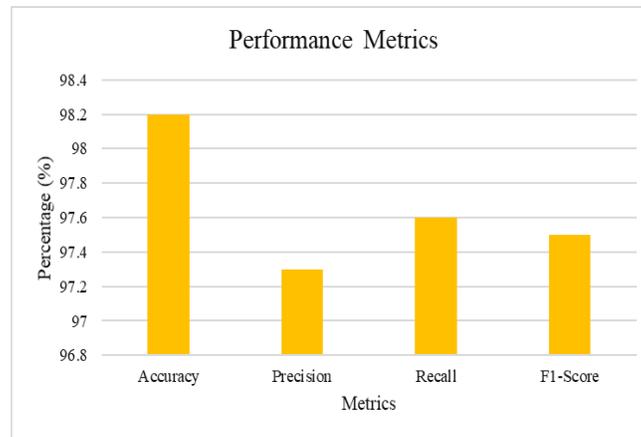


Fig. 2. Performance Metrics of the Proposed Study

The Fig. 2 depicts the GPT-powered formative assessment system's performance indicators. The model obtains an accuracy of 98.2%, suggesting that the predictions are credible. Precision of 97.3% indicates excellent relevance in feedback, whilst recall of 97.6% indicates efficient recognition of relevant occurrences. The F1-score of 97.5% strikes a mix between precision and recall, indicating the model's strong performance in providing individualized, adaptive language learning evaluations.

TABLE II. PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT METHODS

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Rule-based [4]	85.6	83.2	84.5	83.8
Support Vector Machine [13]	92.5	90.0	91.2	90.6
Random Forest [14]	93.7	91.5	92.1	91.8
Naive Bayes [15]	88.3	86.0	87.5	86.7
GPT (Proposed)	98.2	97.3	97.6	97.5

The table II shows that there's a superiority of GPT-based formative assessment compared to traditional and ML strategies. GPT outperforms the remaining two approaches in all of the evaluation metrics, scoring at 98.2% accuracy, providing correct predictions every time. Precision and recall at 97.3% and 97.6%, respectively, indicate that this model can provide relevant and comprehensive feedback with minimal false positives and missing important instances. The 97.5% F1-score reflects excellent balance between precision and recall, guaranteeing optimal performance. Traditional rule-based methods or other ML models are effective in their case but are not at all comparable with the overall accuracy and reliability of GPT.

E. Discussion

The results indicate that the GPT model clearly outperforms in personalizing formative assessments of language learners. Higher metrics rates make GPT extremely effective in the generation of relevant, comprehensive feedback. These findings indicate that GPT can considerably complement mobile learning applications by offering real-time, adaptive support for continuous development in languages. The model is also promising towards future educational use as it strikes a balance between precision and recall.

5. Conclusion and Future Works

This study proposed a new GPT-oriented formative assessment model for mobile-based English language learning and the areas of the deficiencies of the traditional assessment techniques. Riding on the advanced natural language processing prowess of GPT, the suggested model managed to get real-time, personalized feedback which improved learner engagement and increased language acquisition by a great margin. Integration of this model into mobile applications allowed adaptive learning paths, exceeding traditional rule engine and machine learning approach with outstanding accuracy of 98.2%. This innovative approach changes the face of the mobile learning by context-aware immediate feedback, which is consistent with individual learning progress.

Future studies will aim at scaling the GPT-based formative assessment model to cater for diverse linguistic settings and multilingual learning settings. In addition, implementing speech recognition for current spoken reviews and researching inter-platform compatibility in mobile learning applications will be given top priority. Testing the effectiveness of the model in the real-world classroom settings and the effects it has on long-term language retention will add further credence to its practical applicability. Finally, implementation of users' feedback to improve personalization and to perfect adaptive learning mechanisms will contribute to the more appealing and effective m-based language studying.

References

- [1] M. Hasan, A. S. Islam, and I. J. Shuchi, "Using mobile-based formative assessment in ESL/EFL speaking," *J. Lang. Lang. Teach.*, vol. 9, no. 1, pp. 117–125, 2021.
- [2] S. Winarno and H. Al Azies, "The Effectiveness of Continuous Formative Assessment in Hybrid Learning Models: An Empirical Analysis in Higher Education Institutions," *Int. J. Pedagogy Teach. Educ.*, vol. 8, no. 1, pp. 1–11.
- [3] B. Yassin and M. Abdulgalil Abugohar, "Gamified mobile-assisted formative assessment for reviving undergraduate learners' overall language proficiency: a quasi-experimental study," *Teach. Engl. Technol.*, vol. 22, no. 2, pp. 69–89, 2022.
- [4] J. A. Muñoz Velasco and C. González Serrano, "Towards an adaptive Formative Assessment Model in context-based Mobile Learning Systems," *Ing. Compet.*, vol. 26, no. 2, 2024.
- [5] T. R. B. Elbasyouny, "Enhancing students' learning and engagement through formative assessment using online learning tools," Master's Thesis, The British University in Dubai, 2021.
- [6] A. Andujar, "Mobile-mediated dynamic assessment: A new perspective for second language development," *ReCALL*, vol. 32, no. 2, pp. 178–194, 2020.
- [7] N. Yarahmadzehl and M. Goodarzi, "Investigating the role of formative mobile based assessment in vocabulary learning of pre-intermediate EFL learners in comparison with paper based assessment," *Turk. Online J. Distance Educ.*, vol. 21, no. 1, pp. 181–196, 2020.
- [8] J. Wang, B. Cao, P. Yu, L. Sun, W. Bao, and X. Zhu, "Deep learning towards mobile applications," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, IEEE, 2018, pp. 1385–1393.
- [9] S. Fu, H. Gu, and B. Yang, "The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China," *Br. J. Educ. Technol.*, vol. 51, no. 5, pp. 1674–1692, 2020.
- [10] J.-C. Lee and L. Xiong, "Exploring learners' continuous usage decisions regarding mobile-assisted language learning applications: A social support theory perspective," *Educ. Inf. Technol.*, vol. 28, no. 12, pp. 16743–16769, 2023.
- [11] J. Guo, S. Lee, H. Guo, and J. Yang, "Exploring Deep Learning of English Vocabulary Acquisition from Mobile Learning Environment," *J. E-Learn. Res.*, vol. 1, no. 4, pp. 1–16, 2021.
- [12] "English Language Learning." Accessed: Nov. 22, 2024. [Online]. Available: <https://kaggle.com/code/mohdmuttalib/english-language-learning>
- [13] Y. Jin, "[Retracted] Sentiment Analysis in British and American Literature Teaching under Formative Assessment and Machine Learning," *Wirel. Commun. Mob. Comput.*, vol. 2022, no. 1, p. 7954166, 2022.
- [14] Q. Xuan, A. Cheung, and D. Sun, "The effectiveness of formative assessment for enhancing reading achievement in K-12 classrooms: A meta-analysis," *Front. Psychol.*, vol. 13, p. 990196, 2022.
- [15] A. K. Veerasamy, M.-J. Laakso, D. D'Souza, and T. Salakoski, "Predictive models as early warning systems: A Bayesian classification model to identify at-risk students of Programming," in *Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 2*, Springer, 2021, pp. 174–195.