**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

# Planting Seeds of Understanding: Nurturing Listening Comprehension through Formative Assessment and AI-Powered Feedback

**Mohamad Ali Saeedirad [1*], Masoud Khalili Sabet[1]**

[1]Department of English, University of Guilan, Rasht, Guilan, Iran

Email: sabetma2002@yahoo.com

[*]Corresponding author's Email: Saeedirad83@gmail.com

Received: 16-08-2024, Accepted: 14-02-2025

**ABSTRACT**

***This study investigated the effectiveness of AI-powered formative assessments, specifically using ChatGPT, to enhance listening comprehension among intermediate English learners. Sixty participants from a language institute in Rasht, Iran, were randomly assigned to either an experimental group that received AI-driven assessments integrated with continuous, personalized feedback, or a control group that underwent traditional summative assessments of learning to evaluate student learning after the instructional period. Both groups engaged with identical instructional content over 12 weeks, Data were collected through pre-tests, post-tests, and delayed post-tests to assess listening comprehension outcomes. Additionally, participants completed questionnaires and participated in semi-structured interviews to gather feedback on their experiences. The data analysis included paired t-tests to compare performance between groups and thematic analysis for the qualitative data from the surveys and interviews, allowing for a comprehensive evaluation of the effectiveness of AI-enhanced formative assessment. The results revealed that the experimental group significantly outperformed the control group in both immediate and delayed post-tests, underscoring the benefits of AI-driven feedback on improving short-term performance and fostering long-term retention. Additionally, participants in the experimental group reported higher levels of engagement, self-confidence, and autonomy. The study implies that AI-powered formative assessments can improve listening comprehension by offering instant, personalized feedback, but human support is essential to meet learners' emotional and motivational needs.***

**KEYWORDS:** AI-Powered Feedback; Formative assessment; Listening skill; Summative assessment

## INTRODUCTION

Listening comprehension is a vital skill in language learning, serving as the foundation for understanding the rhythm, structure, and nuances of a new language. However, many learners find it especially challenging. This difficulty arises from the complex cognitive processes required, such as decoding sounds, interpreting meanings, and applying contextual knowledge (Zhang & Shen, 2023).

Traditional methods of evaluating listening skills, such as final exams or summative assessments, are useful for measuring overall achievement but don't capture the continuous and evolving nature of the learning process (Dolin et al., 2018). Moreover; the multifaceted nature of listening comprehension, which involves both bottom-up and top-down processing, adds to the complexity of assessment. Bottom-up processing refers to the decoding of linguistic elements, such as sounds, words, and sentences, while top-down processing involves the use of background knowledge, context, and inferencing to construct meaning (Field, 2010). Effective listening requires the combining of these processes, where listeners actively derive meaning by integrating both linguistic and contextual signals. Traditional summative assessments may not adequately capture the nuances of these processes or provide the necessary support for learners to improve their listening abilities.

Formative assessment is characterized by its ongoing, iterative nature, providing continuous feedback that allows for real-time adjustments to the learning process. Unlike summative assessment, which evaluates the result, formative assessment is embedded within the learning journey, offering opportunities for learners to identify strengths, address weaknesses, and refine their understanding continuously (Chen & Zhang, 2019; Fulcher, 2015; Kusumawati, 2020; Ngo, 2019; Tanewong, 2019). Incorporating AI tools like ChatGPT introduces a novel dimension to formative assessment. ChatGPT can simulate real-life conversational contexts, provide immediate feedback, and adapt tasks to the learner's needs, thereby making formative assessment more interactive, personalized, and responsive.

Feedback, as a critical component of formative assessment, is essential for effective learning. Hattie and Timperley (2007) emphasize that feedback, when delivered effectively, can significantly enhance learning and achievement. The integration of AI tools like ChatGPT enhances this feedback loop by offering timely, targeted, and tailored feedback (Parade & Pearl, 2024). In the context of listening comprehension, AI-driven feedback can help learners develop essential metacognitive strategies, such as monitoring their comprehension, predicting content, and drawing inferences, which are crucial for successful listening (Vandergrift & Goh, 2012).

AI-enhanced formative assessment, with its emphasis on continuous feedback and interactive engagement, offers a promising approach to addressing these challenges. By providing ongoing assessment and feedback, AI tools like ChatGPT can help learners develop and refine their listening skills more effectively ( Liu et al. 2024). This study aimed to investigate the impact of AI-enhanced formative assessment on listening comprehension by comparing an experimental group, which received formative assessments with AI-powered feedback, to a control group, which only underwent summative evaluation.

The significance of the present study extended beyond its potential to enhance listening comprehension. It contributed to the broader field of language education by providing evidence-based insights into the integration of AI tools in formative assessment practices. As the global demand for effective language learning continues to grow, there is a pressing need for innovative approaches that support the development of critical language skills (Council of Europe, 2001). This study sought to provide empirical evidence on the effectiveness of AI-enhanced formative assessment in improving listening comprehension, offering valuable insights for educators, curriculum designers, and policymakers. Moreover, it also explored the long-term impact of AI-powered formative assessment on learning retention. Retaining language skills over time is a key challenge in language learning. Traditional summative assessments, while useful for measuring immediate outcomes, do not provide insights into the sustainability of learning gains. By incorporating delayed post-tests and continuous AI feedback, this study evaluated the degree to which formative assessment, supported by AI, could promote enduring improvements in listening comprehension.

This study aimed to investigate the impact of AI-enhanced formative assessment on listening comprehension. The objectives were threefold: first, it aimed to compare the immediate and delayed listening comprehension outcomes of students who received AI-powered formative assessment with those who undertook summative assessment; second, it explored the role of AI-driven feedback in enhancing listening comprehension; and third, it examined the long-term retention of listening comprehension skills. An experimental design with control and experimental groups was

employed, with the experimental group receiving continuous AI-generated feedback, while the control group participated in summative assessment.

By incorporating delayed post-tests and continuous AI feedback, this study also aimed to assess the sustainability of learning improvements. Despite substantial research on formative assessment, there remained a gap in understanding its specific impact on listening comprehension when integrated with AI tools like ChatGPT, particularly concerning long-term retention. This study sought to fill this gap by investigating the effects of AI-driven formative assessment on both immediate and delayed listening comprehension outcomes.

## REVIEW OF RELATED LITERATURE

### THEORETICAL UNDERPINNINGS

Assessment sould be considered as a complex and important task for teachers, a critical component of their teaching, and a foundation of their educational expertise (Edwards, 2017; Moss, 2013). Summative assessments are typically conducted at the end of an instructional period to evaluate the results of student learning. These assessments are high-stakes for everyone involved, especially for students who are being evaluated, and also for their potential use in driving course improvements, assessing teaching effectiveness, and conducting program-level evaluations such as accreditation. In contrast, formative assessments aim to enhance the learning process by providing constructive, nonjudgmental feedback. (Ali, 2024). They focus more on supporting learning rather than judging it. However, assessments often fall on a spectrum between summative and formative, such as when assignments or quizzes in a course contribute to a grade. Therefore, the distinction between summative and formative assessments can vary, with the primary goal influencing the type of assessment, although providing feedback to learners should always be a key element (Kibble, 2017). Final grades involve high-stakes summative judgments, as they can impact students' access to higher education opportunities and influence their career paths (Yildirim et al., 2024), as such**,** the primary purpose of summative assessment is to measure and document student achievements, often to assign grades or evaluate the effectiveness of instruction (Harlen, 2005).

While summative assessment provides valuable information about students' final achievements, it has been criticized for its limitations in informing instruction and supporting ongoing learning. Shepard (2006) argues that summative assessments, by focusing on the end product, often fail to capture the learning process and do not provide the feedback necessary for learners to improve. This can lead to a focus on "teaching to the test," where instruction is narrowly focused on preparing students for summative assessments, potentially at the expense of deeper learning and critical thinking skills (Shepard 2000). Despite these limitations, summative assessment remains an essential component of educational systems worldwide, providing benchmarks for student performance and guiding decisions about curriculum and instruction (Harlen, 2005). However, the integration of AI tools into formative assessment practices offers a complementary approach that can address some of these limitations by providing continuous, personalized feedback and opportunities for improvement throughout the learning process (Harlen, 2005).

Formative assessment, on the other hand, also known as assessment for learning, plays a pivotal role in the educational process by providing continuous feedback that helps learners improve their performance in real-time (Black & Wiliam, 1998). It is grounded in the constructivist theory of learning, which posits that learners actively construct knowledge through interaction with their environment and through reflection on their experiences (Piaget, 1971). This dynamic process enables learners to adjust their understanding and approach based on the feedback they receive, making formative assessment an integral part of the learning journey (Black & Wiliam, 1998).

**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

Formative assessment strategies include a wide range of practices, such as questioning, self-assessment, peer assessment, and the use of formative feedback. These strategies are designed to engage students actively in their learning process, helping them understand their progress and identify areas for improvement (Sadler, 1989). The ongoing nature of formative assessment contrasts with summative assessment, which typically occurs at the end of an instructional period and is used to evaluate student learning outcomes (Nicol & Macfarlane-Dick, 2006;Shute, 2008 ).

## LISTENING COMPREHENSION

Listening comprehension is a crucial component of language acquisition, involving the ability to understand and interpret spoken language (Caruso et al., 2017; Sabet & Mahsefat, 2012). It is a complex, multifaceted skill that requires the integration of various cognitive and linguistic processes. These include phonemic awareness, vocabulary knowledge, syntactic parsing, and the ability to make inferences and understand context (Rost, 2011, Sabet & Mahsefat, 2012 ). Effective listening comprehension involves both bottom-up and top-down processing. Bottom-up processing refers to the decoding of linguistic elements such as sounds, words, and sentences, while top-down processing involves the use of background knowledge, context, and inferencing to construct meaning (Bozorgian & Shamsi, 2023, Vandergrift & Goh, 2012). However, enhancing listening comprehension presents challenges, particularly when traditional assessment methods fail to capture the dynamic and ongoing nature of learning (Dolin et al., 2018). Recent advancements in educational technology, including Artificial Intelligence (AI) tools, have shown significant potential in addressing these challenges by offering personalized and interactive learning experiences that can improve both immediate and long-term listening skills.

One promising approach is the integration of AI-powered tools in self-study modes, such as podcast-based activities. Chaikovska et al. (2020) highlighted the role of AI tools in enhancing the listening comprehension skills of engineering students through self-study podcast activities. In their study, the use of TED podcasts in conjunction with AI-based assessment platforms significantly improved students' listening skills. The experimental group, which engaged with AI-driven podcast assessments, outperformed the control group, supporting the claim that AI tools can create more engaging and effective self-study experiences for students (Chaikovska et al., 2020). This aligned with the current study's hypothesis that AI-enhanced formative assessments could foster deeper engagement and more efficient learning in ESP contexts.

Moreover, research by Suryana et al. (2020) suggested that AI mobile applications, such as Netflix and Tune In, have been found to enhance English listening skills by providing real-time, adaptive feedback and fostering an engaging learning environment. Their findings indicated that students, particularly non-English majors, found these AI tools to be highly effective in improving listening comprehension, largely due to their ability to offer interactive and personalized learning experiences. This further reinforced the argument that AI, when integrated effectively, could lead to significant improvements in listening comprehension (Suryana et al., 2020).

The impact of AI extends beyond the classroom, as intelligent personal assistants (IPAs) like Google Assistant have been shown to promote listening comprehension through interactive, multimodal responses. In a study by Tai and Chen (2024), adolescents using Google Nest Hub with IPA responses demonstrated substantial improvements in listening skills due to the tool's interactive nature, which encouraged active participation and collaboration. The results emphasized that the multimodal feedback provided by IPAs, combined with their ability to adapt to students' needs, can enhance both engagement and cognitive processing, crucial components of effective listening comprehension (Tai & Chen, 2024).

In the context of computer-based language learning, Li (2024) examined the effectiveness of help options in facilitating listening comprehension. Using a meta-analysis of 33 studies, Li found that help options, such as captions

and feedback, significantly enhance L2 listening comprehension, particularly when integrated with computer-based learning tools. This aligns with the findings of the current study, which proposed that AI-powered formative assessments, with their continuous feedback loops, might support sustained improvement in listening comprehension over time. Collectively, these studies provided a compelling body of evidence that AI tools—whether in the form of podcasts, mobile applications, IPAs, or computer-based help options—play a significant role in enhancing listening comprehension skills. By providing personalized, interactive, and adaptive learning experiences, AI may complement traditional teaching methods and foster long-term cognitive growth. The current study built upon this foundation by investigating how AI-driven formative assessments could be integrated into ESP courses to enhance students' listening comprehension skills, particularly within self-study contexts.

The advent of AI technology, such as ChatGPT, has opened new possibilities for enhancing formative assessment practices. AI tools may provide instant feedback, simulate authentic conversational practice, and adapt tasks to meet individual learner needs, thereby offering a more personalized learning experience (Zou et al., 2023). This adaptability is particularly beneficial in the context of language learning, where learners often require tailored feedback to address their unique challenges. Some more studies have reinforced the effectiveness of formative assessment in enhancing student learning outcomes. For example, Bennett (2023) and Brown (2022) found that formative assessment significantly improved student achievement across various subjects, including language learning. The study highlighted the role of timely feedback in helping students develop a deeper understanding of the material, leading to improved performance. Similarly, a study by Brown (2022) demonstrated the positive impact of formative assessment on language learners' motivation and engagement, which in turn, contributed to better learning outcomes.

## FEEDBACK IN LISTENING COMPREHENSION

Feedback is moreover, a critical component of effective language learning, particularly in the development of listening comprehension skills. Feedback provides learners with information about their performance, helping them identify areas for improvement and guiding their learning process (Nicol & Macfarlane-Dick, 2006). In the context of listening comprehension, feedback could take various forms, including corrective feedback on errors, explanations of misunderstandings, and guidance on effective listening strategies.

The effectiveness of feedback in listening comprehension has been the subject of extensive research. For example, a study by Li (2023) found that explicit corrective feedback, where learners were directly informed of their errors and provided with the correct response, was particularly effective in improving listening comprehension. It also highlighted the benefits of interactive feedback, where learners were encouraged to reflect on their listening strategies and considered alternative approaches.

Recent advances in AI technology have further enhanced the potential of feedback in language learning. AI-powered tools, such as ChatGPT, might provide instant feedback on listening comprehension tasks, allowing learners to immediately address their errors and refine their understanding (Liu et al., 2024). AI tools could also offer personalized feedback if they were tailored to the learner's specific needs, thereby supporting more targeted and effective learning. Another study by Zou et al. (2023) demonstrated the benefits of AI-generated feedback in listening comprehension, showing that learners who received personalized feedback from an AI tool made significant gains in their listening skills compared to those who received traditional feedback (Zou et al., 2023).

According to the recent studies the integration of AI tools into formative assessment practices represents a promising approach to enhancing listening comprehension. By providing continuous, personalized feedback and opportunities for practice, AI-powered formative assessment can help learners develop and refine their listening skills more effectively than traditional methods (Zhang, et al. 2023). This study by investigating the impact of AI-enhanced

formative assessment on listening comprehension, contributed to the growing body of research on the use of AI in language education and offered practical insights for educators and curriculum designers.

## CHATGPT AND AI IN LANGUAGE LEARNING

AI-powered tools, such as ChatGPT, have shown great potential in revolutionizing language learning by offering personalized, interactive, and adaptive experiences. As a large language model developed by OpenAI, ChatGPT engages learners in natural language conversations, simulates real-world communication scenarios, and provides instant feedback on language use (Brown et al., 2020). This interactivity and adaptability would make ChatGPT a particularly valuable tool for enhancing formative assessment practices, which are designed to offer continuous feedback and to support the learning process.

More recent studies have highlighted the effectiveness of ChatGPT in language learning. For instance, Xu et al. (2024) explored the use of ChatGPT in language instruction and found that students who engaged with AI-powered tools showed significant improvements in their language skills, including listening comprehension. It emphasized how personalized feedback and interactive practice play a crucial role in fostering language development. In a similar vein, Liu et al. (2024) demonstrated that AI-powered feedback notably enhanced language learning outcomes by providing learners with timely, relevant, and personalized responses, which allowed them to identify and address their weaknesses. These findings underline the potential of incorporating ChatGPT into formative assessment practices to provide continuous, tailored feedback that enhances listening comprehension.

ChatGPT's ability to continuously monitor student progress and provide targeted feedback makes it a powerful tool for improving learners' engagement, motivation, and ultimately, their listening comprehension skills. As AI technology continues to evolve, it is expected that tools like ChatGPT would play an increasingly important role in language education. They offer new possibilities for creating more personalized, interactive, and adaptive learning experiences, thereby reshaping how listening comprehension is taught and assessed.

## RESEARCH QUESTIONS

The present study purported to investigate the impact of AI-powered formative feedback on listening comprehension outcomes. Specifically, the following research questions guided the study:

1. How does AI-powered formative feedback affect learners' listening comprehension performance as measured by pretests, posttests, and delayed posttests?
2. How does AI-powered feedback influence the consistency of learners' performance (e.g., standard deviations) compared to traditional feedback methods?
3. What challenges do learners face when interacting with AI-powered feedback systems in the context of listening comprehension?
4. To what extent does AI-powered feedback address learners' emotional and motivational needs, and how does it compare to human-provided feedback?

## METHODOLOGY

The study employed an experimental design with control and experimental groups to investigate the impact of AI-enhanced formative assessment on listening comprehension. The study aimed to compare the immediate and delayed listening comprehension outcomes of students who received AI-enhanced formative assessment with those who only

undertook summative assessment. Additionally, the study sought to explore the role of AI-powered feedback in enhancing listening comprehension within the formative assessment framework.

## PARTICIPANTS

The study involved 60 language learners enrolled in an intermediate-level English language course at a language institute in Rasht, Iran. The participants were randomly assigned to either the experimental group (n=30) or the control group (n=30). The experimental group received AI-enhanced formative assessment, while the control group underwent traditional summative assessment. All participants took a pretest to ensure homogeneity in their listening comprehension skills before the intervention.

## INSTRUMENTS

The study utilized a variety of instruments to assess listening comprehension and gather data on the participants' learning experiences. These included:

- Pre-test and Post-test: Standardized listening comprehension tests are administered at the beginning (pretest) and end (post-test) of the study to measure the participants' listening comprehension skills. The tests included multiple-choice questions, gap-fill exercises, and short-answer questions based on authentic listening passages.

- Delayed Post-test: A delayed post-test was administered six weeks after the intervention to assess the long-term retention of listening comprehension skills. The delayed post-test was similar in format to the pre-test and post-test.

- Questionnaires: Participants completed questionnaires at the end of the study to provide feedback on their learning experiences, including their perceptions of the AI-enhanced formative assessment and the feedback provided by ChatGPT.

- Interviews: Semi-structured interviews were conducted with a subset of participants from the experimental group to gain deeper insights into their experiences with the AI-enhanced formative assessment and the impact of ChatGPT on their listening comprehension skills.

## PROCEDURE

The study was conducted over a 12-week period, during which both the experimental and control groups participated in a series of listening comprehension lessons. The lessons covered a range of topics and listening tasks, including conversations, lectures, and news reports. The instructional content was identical for both groups, but the assessment methods differed.

## EXPERIMENTAL GROUP: AI-ENHANCED FORMATIVE ASSESSMENT

Participants in the experimental group received formative assessment throughout the course, with continuous feedback provided by ChatGPT. The AI tool was integrated into the learning management system (LMS) used by the university, allowing learners to interact with ChatGPT during listening tasks. ChatGPT delivered a range of listening exercises, including comprehension questions, summary tasks, and critical thinking prompts based on the listening passages. After completing each exercise, learners receive immediate, personalized feedback from ChatGPT, which highlights areas of strength and provided suggestions for improvement.

In addition to providing feedback on specific tasks, ChatGPT offered ongoing support through regular check-ins, where learners could ask questions, clarify misunderstandings, and seek guidance on effective listening strategies. The AI tool also tracked each learner's progress and adapted the difficulty of the listening tasks based on their

**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

performance, ensuring that the exercises remained challenging yet achievable. The formative assessment process was iterative, with learners encouraged to revisit tasks and apply the feedback provided by ChatGPT. This continuous cycle of assessment and feedback aimed to promote deeper learning and retention of listening comprehension skills.

## CONTROL GROUP: TRADITIONAL SUMMATIVE ASSESSMENT

Participants in the control group followed the same instructional content as the experimental group but were assessed only through summative methods. They completed a single summative test at the end of the course, which included a listening comprehension exam similar to the pretest and posttest used in the study. Unlike the experimental group, the control group did not receive continuous feedback or formative assessment during the course. This study aimed to explore the impact of AI-enhanced formative assessment on listening comprehension, contributing to the growing body of research on the use of AI in language education and offering practical insights for educators and curriculum designers. The findings of this study were expected to provide valuable insights into the effectiveness of AI-enhanced formative assessment in promoting listening comprehension and long-term retention of language skills. Furthermore, the study contributed to the ongoing discourse on the role of AI in education, particularly in the context of language learning, where personalized and adaptive feedback was critical for success.

## DATA ANALYSIS

The study assessed two groups, the Experimental Group (Group A) and the Control Group (Group B), at two different points in time: immediately after the treatment (T1-T3) and after a four-week interval (T4-T6) to evaluate the long-term retention effects of the formative assessment approaches used (Harris, 1969). To prevent any potential bias in the results, participants were not informed about the delayed post-test, ensuring that no additional practice influenced the outcomes.

Additionally, an eight-item Likert scale survey was administered to Group A to gauge their perceptions of the AI-powered formative assessment process. The survey results are discussed in the final section of this analysis.

Table 1 and Table 2 present the performance data for Groups A and B across all tests, including the mean, standard deviation, maximum and minimum scores, standard error of the mean, and variance for each test. The subsequent discussion compares these scores in detail.

**Table1**
*Representative data of the performance of Group A in all TESTS.*

| **Time** | **Post-test** | | | **Delayed Post-test** | | |
|---|---|---|---|---|---|---|
| **Test** | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ |
| **Mean** | 17.433 | 10.766 | 6.666 | 17.4667 | 10.566 | 7.066 |
| **Std. Deviation** | 5.468 | 3.588 | 2.682 | 3.4813 | 2.329 | 1.837 |
| **Maximum** | 25.00 | 19 | 10.00 | 25.00 | 15.00 | 10.00 |
| ***Minimum*** | 8.00 | 3.00 | 1.00 | 11.00 | 7.00 | 2.00 |
| ***Std. Error*** | .998 | .6551 | .4897 | .6356 | .425 | .335 |
| ***Variance*** | 29.909 | 12.875 | 7.195 | 12.120 | 5.426 | 3.375 |

**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

**Table2**
*Representative data of the performance of Group B on all TESTS.*

| **Time** | **Post-test** | | | **Delayed Post-test** | | |
|---|---|---|---|---|---|---|
| **Test** | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ |
| **Mean** | 12.466 | 9.433 | 5.800 | 11.60 | 9.133 | 5.366 |
| **Std. Deviation** | 4.066 | 2.344 | 2.074 | 3.644 | 2.129 | 1.771 |
| **Maximum** | 23.00 | 15.00 | 9.00 | 19.00 | 13.00 | 8.00 |
| **Minimum** | 5.00 | 3.00 | 2.00 | 5.00 | 4.00 | 2.00 |
| **Std. Error mean** | .7423 | .4280 | .3787 | .665 | .388 | .3233 |
| **Variance** | 16.533 | 5.495 | 4.303 | 13.28 | 4.533 | 3.137 |

Comparison and T-Test Analysis:

Using the t-test formula, the t value was calculated based on the data from Tables 1 and 2. Table 3 shows the matched pair *t value* between the two groups, indicating significant differences in their performance.

**Table3**
*Matched pair t value of two groups*

| **Groups** | **Mean** | **Std. Deviation** | **Std. Error Mean** | **95% Confidence Interval of the Difference** | | ***t* value** | **df. N-1** | **Sig. (2 tailed)** |
|---|---|---|---|---|---|---|---|---|
| | | | | **Lower** | **Upper** | | | |
| **A & B** | 4.966 | 8.176 | 1.492 | 1.913 | 8.019 | ***3.327*** | 29 | ***2.045*** |

The *t value* (3.327) is significantly higher than the critical value (2.045), indicating a statistically significant difference between the performances of the two groups. The Experimental Group (Group A) outperformed the Control Group (Group B) on both immediate and delayed post-tests, suggesting that the AI-powered formative assessment tools had a positive impact on their learning outcomes.

**Table 4**
*Differences and the subtraction of the scores between groups A & B in $T_2$ & $T_3$*

| **$T_2$** | | | | **$T_3$** | | | |
|---|---|---|---|---|---|---|---|
| **Statistics** | A | B | A-B=X | **Statistics** | A | B | A-B=X |
| **Mean** | 10.433 | 9.433 | 1 | **Mean** | 6.666 | 5.800 | 0.866 |
| **Std. deviation** | 3.244 | 2.344 | 0.900 | **Std. deviation** | 2.6822 | 2.074 | 0.607 |
| **Std. Error mean** | .592 | .4280 | 0.164 | **Std. Error mean** | .4897 | .378 | 0.110 |
| **Variance** | 12.87 | 5.495 | 7.38 | **Variance** | 7.195 | 4.303 | 2.892 |

**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

Table 4 illustrates the differences in scores between Groups A and B in the immediate multiple-choice test (T2) and cloze test (T3). The positive difference in mean scores indicates that Group A performed better than Group B in these assessments, likely due to the effectiveness of the AI-powered feedback.

**Table 5**
*Matched pair t-value.*

| Group | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | **t value** | df. N-1 | Sig. 2 tailed |
|---|---|---|---|---|---|---|---|---|
| | | | | Lower | Upper | | | |
| A & B | 5.866 | 4.861 | .887 | 4.051 | 7.682 | **6.609** | 29 | **2.045** |

The same pattern is observed in Table 5*, which shows the t value for the delayed post-tests (T4). Here, the t value (6.609) is even more significant, reinforcing the earlier findings that AI-powered formative assessment tools have a lasting positive effect on student performance.

**Table 6**
*Differences and the subtraction of the scores between groups A & B in $T_5$ & $T_6$.*

| | **$T_5$** | | | | **$T_6$** | | |
|---|---|---|---|---|---|---|---|
| **Statistics** | A | B | A-B=X | **Statistics** | A | B. | A-B=X |
| **Mean** | 10.566 | 9.133 | 1.433 | **Mean** | 7.066 | 5.366 | 1.7 |
| **Std. deviation** | 2.329 | 2.129 | 0.200 | **Std. deviation** | 1.837 | 1.771 | 0.900 |
| **Std. Error mean** | .425 | .388 | 0.036 | **Std. Error mean** | .335 | .3233 | 0.164 |
| **Variance** | 5.426 | 4.533 | 0.893 | **Variance** | 3.375 | 3.137 | 0.238 |

In both delayed tests, Group A continued to outperform Group B, suggesting that the AI-powered tools also helped in the long-term retention of the material.

## PARTICIPANT ATTITUDES

After the treatment and tests, a survey was conducted to evaluate the attitudes of Group A towards the AI-powered formative assessment process. The results from the survey were overwhelmingly positive, as evidenced by the high percentage of learners who expressed satisfaction with various aspects of the process.

**Journal of Applied Linguistics Studies, Vol.4, No.2, 2024: 34-50**
*https://jals.aliabad.iau.ir*
ISSN: 2820-9974

| No. | Category | Percentage | Interpretation |
|---|---|---|---|
| 1 | Found using AI tools for practicing listening interesting | 97.2% | A vast majority found AI tools engaging for listening practice. |
| 2 | Preferred AI-powered feedback over traditional methods | 98.6% | Almost all learners favored AI feedback over traditional methods. |
| 3 | Found the technology easy to use | 87.2% | Most learners found the technology user-friendly and accessible. |
| 4 | Felt comfortable doing exercises in a computer-based environment | 84.6% | A large portion felt at ease using computers for their exercises. |
| 5 | Found the units and exercises comprehensible | 92.6% | The majority understood the content and exercises clearly. |
| 6 | Reported studying all lessons | 95.2% | Most learners were committed to studying all lessons provided. |
| 7 | Completed all assigned exercises | 96.6% | Nearly all learners successfully completed their assigned tasks. |
| 8 | Felt they had adequate time to complete the exercises | 88% | Most learners were satisfied with the time available to finish the exercises. |
| 9 | Followed the teacher's recommendations | 99.2% | An overwhelming majority adhered to their teacher’s guidance and recommendations. |

Specifically, 97.2% of participants found using AI tools for practicing listening interesting, indicating strong engagement with the technology. Furthermore, 98.6% preferred AI-powered feedback over traditional methods, highlighting its effectiveness and appeal. Most learners reported a high level of ease with the technology, with 87.2% finding it user-friendly, and 84.6% feeling comfortable completing exercises in a computer-based environment. The comprehensibility of the units and exercises was also highly rated, with 92.6% of participants finding them clear and understandable.

Commitment to the program was strong, with 95.2% of learners studying all lessons and 96.6% completing all assigned exercises. Additionally, 88% felt they had adequate time to complete the exercises, suggesting that the pace of the program was well-managed. Finally, an impressive 99.2% of learners followed the teacher’s recommendations, reflecting a high level of adherence to the guidance provided throughout the process. Overall, these results suggest that Group A had a very positive experience with AI-powered formative assessment, demonstrating both its appeal and effectiveness in enhancing the learning process.

## PARTICIPANTS’ PERCEPTIONS OF AI-POWERED LISTENING COMPREHENSION AND FEEDBACK

This study incorporated a semi-structured interview with twelve participants from the experimental group, aiming to gather their perceptions after a classroom intervention where ChatGPT and other AI tools were employed for formative assessment in listening comprehension. The qualitative analysis of their responses revealed several key themes, which are discussed below. The overall response to using AI-powered tools for listening comprehension was overwhelmingly positive. For many students, this was their first time engaging with AI in a learning context. Ahmad described his

experience as "fun and exciting," emphasizing that this novel approach not only engaged him but also made learning more enjoyable. The novelty of using AI tools like ChatGPT appeared to significantly enhance his learning experience, making it both effective and memorable.

Participants such as Reza highlighted the effectiveness of AI tools in reinforcing learning. He believed that ChatGPT helped him retain information for longer periods, making the learning experience more impactful. This view was echoed by other students who appreciated how the AI facilitated deeper understanding and long-term retention of the material. The AI's ability to make learning "stick" was a recurring theme, suggesting that students found the technology both engaging and beneficial to their educational outcomes. A notable outcome of using AI tools was the increase in students' self-confidence and their sense of independence in learning. Shima shared that practicing listening with ChatGPT allowed her to repeat exercises as many times as she needed, without the embarrassment she previously felt in traditional classroom settings. Before using AI, she had been hesitant to ask teachers to repeat information for multiple times, fearing to be judged by her peers. The AI tool, however, empowered her to take control of her learning, developing a more confident and self-directed approach.

Similarly, Masoumeh reflected on the raise of confidence she experienced, noting the significant advantage of being able to replay content without the fear of being laughed at by classmates. With a sense of humor she mentioned that she could ask the tool to repeat content "a million times" if necessary, underscoring the personalized and pressure-free learning environment that AI tools can provide. Ladan and Amir emphasized the user-friendly nature of AI tools, highlighting the comfort and control these tools offered compared to traditional learning methods. Ladan found the AI tools more inspiring and easier to navigate, while Amir appreciated the immediate feedback and corrections provided by ChatGPT. This instant response capability allowed him to learn at his own pace, free from the pressure that often accompanies teacher-led instruction. The sense of control over their learning process was a significant factor in the positive perceptions of AI tools.

Engagement was another critical theme that emerged from the interviews. Ahmad noted that using ChatGPT made the learning process so enjoyable that he often lost track of time. He particularly appreciated the interactive aspect of AI tools, such as the ability to compare his speaking with that of native speakers. This added a personal challenge to his practice sessions, enhancing his motivation and engagement. Shima added that the AI’s adaptability made the exercises more engaging than traditional methods. She valued the ability to adjust the speed, tone, and accent of the listening materials, tailoring the exercises to her specific needs and proficiency level. This level of personalization allowed her to experiment with different voices and accents, deepening her understanding and making the learning process more dynamic and interesting.

Despite the overall positive experiences, students did encounter some challenges. For example, Akram noted that adjusting to the AI's interface and practicality required initial effort. She expressed a desire for more comprehensive training to better adjust the tool’s full capacities. Moreover, Mina, another participant, mentioned the potential benefits of integrating visual aids, such as pictures or illustrations, to complement textual explanations. Her feedback highlights the need for ongoing support and instruction in using AI tools to their fullest potential. Shahrokh, another participant, provided a contrasting perspective, emphasizing a limitation of AI tools: their inability to understand human emotions and feelings. He pointed out that while AI can provide technical feedback, it cannot sense when a student is unwell or emotionally distressed—something human teachers are naturally more attuned to. This limitation suggests that, while AI tools are valuable, they cannot entirely replace the empathetic support offered by human educators.

When comparing AI tools to traditional methods, students overwhelmingly favored AI for its flexibility and non-intimidating environment. The ability to interact with content at their own pace, free from the pressure of keeping up with the class, was seen as a significant advantage. This individualized approach was particularly appreciated, as it allowed students to engage more deeply with the material and develop their skills in a manner tailored to their

personal learning style. While students were generally satisfied with the AI tools, they did offer suggestions for improvement. Several participants expressed a desire for more interactive features, such as the ability to simulate real-life listening scenarios or engage in conversational practice with the AI. Additionally, there was interest in further customization options, enabling students to tailor the exercises even more closely to their learning styles and preferences. The adaptability of AI feedback was highly valued by the participants. Reza, for instance, appreciated the ability to request repeated explanations until he fully understood the content. This dynamic learning process was seen as a major advantage over traditional methods, which may not allow for such personalized repetition.

Shima and Amir also praised the clarity and adaptability of the feedback. Amir found the ability to adjust accents particularly useful, as it helped him overcome difficulties with British accents by switching to American accents, which were easier for him to comprehend. Zahra added that the AI's capacity to explain unfamiliar words or phrases multiple times, and even paraphrase complex texts into simpler language, made the learning process more accessible and enjoyable. However, not all participants found the AI tools entirely inspirational. Akram's experience highlighted a gap in user proficiency with the technology, indicating a need for better training. She expressed frustration at not being able to fully utilize the tool's capabilities, such as generating visual aids to enhance understanding. This suggests that while AI tools have significant potential, their effectiveness partly depends on the user's ability to navigate and employ them effectively.

Despite some challenges, the overall perception of AI-powered listening comprehension and feedback was highly positive. Students were enthusiastic about the potential of these tools to enhance other areas of language learning, such as reading, writing, and speaking. Nahid, for example, suggested that ChatGPT could be particularly useful for revising writing assignments, noting that AI could provide immediate feedback, whereas feedback from human teachers often takes much longer. This enthusiasm reflects a strong acceptance of technology among students and recognition of its potential to boost and improve the learning process. The ability to receive real-time, personalized feedback was seen as a major benefit, suggesting that AI tools could play a significant role in future educational practices.

## DISCUSSION

This study demonstrated the transformative impact of AI-powered formative assessment tools on immediate and long-term listening comprehension outcomes. The results confirmed the benefits of integrating AI into formative assessments, extending beyond the well-documented advantages of traditional formative assessment practices. The experimental group (Group A), which engaged with AI-driven formative assessments, consistently outperformed the control group (Group B) across all assessments. This superior performance was evident both immediately after the intervention (T1-T3) and persisted after a four-week delay (T4-T6). These findings aligned with the broader educational literature highlighting the role of formative assessment in improving learning outcomes (Bennett, 2023; Black & Wiliam, 1998; & Brown, 2022). The integration of AI tools, such as ChatGPT appeared to enhance these benefits by offering immediate, personalized feedback tailored to individual learning needs, which was crucial for the complex task of listening comprehension (Zou et al., 2023).

Statistical analysis further supported these conclusions, with t-values significantly surpassing the critical value, underscoring the reliability of the observed differences between the groups. This statistical significance reinforced the argument that AI-enhanced formative assessment tools were effective in language education, especially in providing timely and targeted feedback critical for developing listening comprehension skills (Hattie & Timperley, 2007).

Analysis of specific tests revealed that Group A, not only achieved higher mean scores but also exhibited lower standard deviations, compared to Group B. Such a greater consistency in performance suggested that the personalized and adaptive nature of AI feedback contributed to more uniform learning outcomes. By situating the findings of this study within the evolving body of literature that has just reviewed, it is shown that AI-powered feedback

operationalizes and advances these foundational principles. The reduced standard deviations and improved consistency observed in Group A reflected the precision and adaptability of AI tools, which align with the rich body of literature in the field. The enduring impact of AI-driven formative assessment on long-term retention was particularly notable. Group A's continued superior performance in delayed post-tests (T4-T6) highlighted the effectiveness of AI tools like ChatGPT in reinforcing material and promoting sustained retention of listening comprehension skills. This finding supported the broader literature emphasizing the role of formative assessment in enhancing long-term educational outcomes (Dmitrenko et.al, 2021; Vandergrift & Goh, 2012; Harris, 1969).

Student feedback from Group A offered additional insights into the effectiveness of AI tools. The overwhelmingly positive responses reflected high levels of engagement and satisfaction with the AI-powered formative assessment process. This positive attitude likely contributed to the improved performance observed in Group A, as engagement and motivation are critical factors in successful learning (Jurado et.al, 2023). The interactive and responsive nature of AI feedback seemed to make the learning process more enjoyable and less intimidating, in addition, it might enhance motivation and engagement (Xu et al., 2023).

Despite the advantages, the study also identified challenges associated with AI tools. Some participants experienced difficulties with the AI interface, suggesting that the effectiveness of AI-driven formative assessment depended on users' proficiency with the technology. Additionally, the inability of AI to address emotional and motivational aspects of learning highlighted the need for a complementary role for human educators. AI tools should enhance rather than replace human interaction, particularly in areas where emotional support and motivational encouragement are crucial (Shepard, 2000).

In conclusion, this study provided compelling evidence that AI-powered formative assessment tools significantly enhanced listening comprehension skills both immediately and over the long term. The positive feedback from students, combined with substantial performance improvements, underscored the potential of AI to revolutionize formative assessment practices in language education. However, ongoing training and support were essential to completely harness the capabilities of AI tools, and human educators continued to play a vital role in addressing aspects of learning that AI could not replicate.

As educational technologies evolve, the integration of AI which offers more promising opportunities for creating personalized, engaging, and effective learning experiences is required. This study contributed valuable insights to the growing body of research on AI in education, offering practical implications for educators, curriculum designers, and policymakers.

## CONCLUSION AND IMPLICATIONS

The findings of this study highlight the significant impact of AI-powered formative assessment tools on enhancing both immediate and long-term listening comprehension outcomes. The experimental group, which utilized AI-driven assessments, consistently outperformed the control group in both immediate and delayed post-tests, underscoring the effectiveness of AI tools like ChatGPT in providing personalized, timely feedback that fosters deeper learning and retention.

The statistical analyses reinforced these findings, demonstrating the reliability and significance of the observed improvements. The integration of AI-driven formative assessment not only facilitated higher levels of engagement but also contributed to learners' increased self-confidence and independence in language learning. Students in the experimental group expressed satisfaction with the flexibility and immediacy of AI-powered feedback, appreciating how it adapted to their individual learning needs. These results emphasize the transformative potential of AI in language education by offering a dynamic, student-centered learning experience.

Despite its advantages, the study also identified certain challenges associated with AI-powered formative assessment tools. Some learners struggled with navigation due to lower digital literacy, leading to frustration and delays in utilizing feedback effectively. Additionally, while AI systems provide precise, data-driven feedback, they lack the human capacity to recognize learners' emotional states, such as anxiety, frustration, or the need for encouragement. This limitation highlights a crucial gap in AI's ability to provide holistic language learning support. Although AI enhances instructional efficiency and personalized feedback, it cannot fully replace the role of human educators, who offer motivation, emotional support, and contextual understanding that are essential for language acquisition. These findings suggest that AI tools should serve as a supplement rather than a substitute for human instruction, combining AI's analytical strengths with the empathetic and motivational guidance of teachers.

The educational implications of this study emphasize the potential of AI-powered formative assessments to revolutionize listening comprehension instruction by providing immediate, individualized feedback that enhances engagement, motivation, and learning outcomes. However, effective implementation requires adequate training for both students and educators to ensure optimal use of AI tools. Integrating AI into language education should involve a balanced approach that combines technological advancements with human interaction, ensuring that learners receive both the cognitive and emotional support necessary for language development. Future research should further explore the long-term impact of AI-powered formative assessment on other language skills, investigate ways to improve AI's user interface and adaptability, and examine how AI can be integrated with human instructional support to create a more comprehensive and effective language learning environment.

## REFERENCES

Ali, Q. I. (2024). Towards more effective summative assessment in OBE: a new framework integrating direct measurements and technology. *Discover Education, 3*(1), 107.

Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment, 28*(2), 83–104. https://doi.org/10.1080/10627197.2023.2202312

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Bozorgian, H., & Shamsi, E. (2023). A review of research on metacognitive instruction for listening development. *International Journal of Listening*, 1-16. https://doi.org/10.1080/10904018.2023.2197008

Brown, K. (2022). Using formative assessments to motivate students in English language arts (Capstone projects and master's theses, No. 1304). California State University, Monterey Bay. https://digitalcommons.csumb.edu/caps_thes_all/1304

Brown, T. B., Mann, B., & Ryder, N. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Caruso, M., Gadd Colombi, A., & Tebbit, S. (2017). Teaching how to listen: Blended learning for the development and assessment of listening skills in a second language. *Journal of University Teaching & Learning Practice, 14*(1). https://doi.org/10.53761/1.14.1.7

Chaikovska, O., Semenyshena, I., Stoliarenko, O., & Hlushkovetska, N. (2020). Enhancing students' listening comprehension skills through AI-based podcast activities: A study in self-study mode. *JOLLT Journal of Languages and Language Teaching, 8*(1), 48–59. https://doi.org/10.33394/jollt.v8i1.2221

Chen, J., & Zhang, L. J. (2019). Assessing student-writers' self-efficacy beliefs about text revision in EFL writing. *Assessing Writing, 40*(March), 27–41. https://doi.org/10.1016/j.asw.2019.03.002

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Dmitrenko, N., Budas, I., Koliadych, Y., & Poliarush, N. (2021). Impact of formative assessment on students' motivation in foreign language acquisition. *East European Journal of Psycholinguistics, 8*(2), 36-50. https://doi.org/10.29038/eejpl.2021.8.2.dmi

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In J. Dolin & R. Evans (Eds.), *Transforming assessment* (Vol. 4, pp. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3

Edwards, F. (2017). The development of summative assessment literacy: An exploration of the experiences of beginner secondary science teachers in New Zealand. Doctoral Thesis. University of Waikato.

Field, J. (2010). Listening in the language classroom. *ELT Journal, 64*(3), 331-333.

Fulcher, G. (2015). *Re-examining language testing*. Routledge.

Harlen, W. (2005). Teachers' summative practices and assessment for learning–tensions and synergies. *The Curriculum Journal, 16*(2), 207-223.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Jurado, D. B., Guevara, C., Jordan, A. J., & Castillo, D. P. (2023). Auditory skills in English language learning and the use of technological devices as strategies. *Journal of Higher Education Theory and Practice, 23*(17), 236-243.

Kibble, J. D. (2017). Best practices in summative assessment. *Advances in physiology education, 41*(1), 110-119.

Kusumawati, A. J. (2020). Redesigning face-to-face into online learning for speaking competence during COVID-19: ESP for higher education in Indonesia. *International Journal of Language Education, 4*(2), 276–288. https://doi.org/10.26858/ijole.v4i2.14745

Li, H. (2023). A review on corrective feedback research of the recent 20 years. *International Journal of Education and Humanities, 9*(3), 190-195.

Li, R. (2024). Do help options facilitate computer-based L2 listening comprehension? Evidence from an activity theory perspective. *British Educational Research Association*, 46(1), 15–30. https://doi.org/10.1002/ber.21123

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2024.2310288

Moss, C. M. (2013). Research on classroom summative assessment. In J. H. McMillan (Ed.), Sage handbook on research on classroom assessment (pp. 235–255). Sage

Ngo, N. (2019). Understanding the impact of listening strategy instruction on listening strategy use from a socio-cultural perspective. *System, 81*, 63-77. https://doi.org/10.1016/j.system.2019.01.002

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.

Parade, H., & Pearl, J. (2024). The role of ChatGPT in education: Enhancing cognitive skills development through generative AI. *ResearchGate*. https://doi.org/10.13140/RG.2.2.14678.87366

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*.

Rost, M. (2013). *Teaching and researching: Listening*. Routledge.

Sabet, M. K., & Mahsefat, H. (2012). The impact of authentic listening materials on elementary EFL learners' listening skills. *International Journal of Applied Linguistics and English Literature, 1*(4), 216-229. http://doi:10.7575/ijalel.v.1n.4p.216

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 623–646). Praeger.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Suryana, I., Asrianto, & Murwantono, D. (2020). Artificial intelligence to master English listening skills for non-English major students. *Computer Assisted Language Learning, 33(3)*, 276-289. https://doi.org/10.1111/bjet.13532

Tai, T.-Y., & Chen, H. H.-J. (2024). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Language Learning & Technology*, 28(1), 45-61. https://doi.org/10.1016/j.system.2022.102970

Tanewong, S. (2019). Metacognitive pedagogical sequence for less-proficient Thai EFL listeners: A comparative investigation. *RELC Journal, 50*(1), 86–103. https://doi.org/10.1177/0033688218754942

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Xu, X., Su, Y., Zhang, H., Zhang, Y., & Hao, S. (2024). Beyond theory: A mixed-methods investigation of postgraduate engagement with ChatGPT for IELTS speaking. *Research Square, 14(2),* 1-29. https://doi.org/10.21203/rs.3.rs-4503199/v1

Yildirim, A., Stjernkvist, M. P., Hilden, R., Fröjdendahl, B., & Oscarson, A. D. (2024). Developing summative assessment literacy: Novice language teachers' perceptions in Sweden and Finland. *Review of Education*, 12(2), 34-87. DOI: 10.1002/rev3.3487

Zhang, R., Zou, D., & Cheng, G. (2023). A systematic review of technology-enhanced L2 listening development since 2000. *Language Learning & Technology*, *27(3)*, 41-64. https://hdl.handle.net/10125/73531