Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of Human Centered AI Supported Instruction on EFL Listening With Adaptive Feedback Gamified Engagement and Motivational Scaffolding

Aliakbar Tajik

tajik.esl.teacher2024@gmail.com

Islamic Azad University of Varamin

Atefeh Karkhaneh

Islamic Azad University of Varamin

Research Article

Keywords: Gamified Learning Environments, EFL Listening Comprehension, AI-supported language learning systems, Adaptive Feedback, Motivational Scaffolding, Learner Engagement, Self-Efficacy, Tajik's Scaffolded Motivator Model (T-SMM)

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9922866/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Listening comprehension remains one of the most cognitively demanding yet comparatively underemphasized components of English as a Foreign Language (EFL) instruction, particularly in adolescent classrooms. Addressing this gap, the present mixed-methods study investigated the effects of AI-supported instruction on learners' listening comprehension, engagement, and motivation. Two instructional conditions were implemented under controlled classroom settings: a gamified AI-driven personalized learning system (Duolingo) incorporating progression tracking and structured feedback, and a non-gamified conversational AI platform (Replika) designed for personalized dialogue practice without game-based reward mechanisms such as points or badges. The study was conceptually informed by Schema Theory, Self-Regulation Theory, Dynamic Assessment within Sociocultural Theory, and Gamification Theory, which were synthesized into two integrative frameworks—the AI-Enhanced Language Learning Matrix (AELLM) and the Intelligent Learning Environment Design (ILED). Based on the mixed-methods findings, Tajik's Scaffolded Motivator Model (T-SMM) was proposed as an explanatory model of motivational scaffolding in AI-mediated listening instruction. Participants included 53 Iranian high school learners (aged 14–15) who completed a 12-week instructional program comprising 24 sessions. Quantitative results indicated significant improvements in listening proficiency and self-efficacy in the gamified condition, while qualitative findings revealed enhanced engagement and perceived instructional support. The findings suggest that structured feedback and gamified scaffolding may support sustained participation and deeper cognitive engagement. These results contribute to the design and evaluation of AI-supported EFL listening environments and offer practical implications for technology-enhanced language instruction.

1. Introduction

Listening comprehension constitutes one of the most cognitively demanding pillars of second language acquisition, accounting for roughly 50–80% of total communicative engagement in language learning (Yıldırım & Yıldırım, 2016; Hosseini et al., 2021). As a multidimensional cognitive process, it involves the rapid orchestration of phonological decoding, lexical access, syntactic parsing, and semantic integration to construct coherent meaning from incoming auditory input (Buck, 2001; Goh & Vandergrift, 2021). In real-time communication, learners must also cope with extralinguistic variations—such as accent, speech rate, and discourse context—that further complicate comprehension (Rost, 2016). Academic and professional environments intensify these challenges, exposing learners to diverse discourse types ranging from informal dialogue to discipline-specific lectures, each demanding distinct adaptive strategies (Field, 2010). Consequently, there is a growing pedagogical need for instructional approaches that support both linguistic decoding and learners' ability to adapt to authentic communicative conditions.

Recent developments in digital and AI-supported learning environments have created new opportunities to enrich listening instruction by providing repeated exposure to auditory input, interactive practice, and immediate feedback. Such tools can complement classroom instruction by offering structured listening

activities and opportunities for individualized practice outside the traditional classroom setting. Within this evolving landscape, educational platforms differ substantially in their instructional design, particularly in the extent to which they incorporate gamified progression systems, structured feedback mechanisms, and conversational interaction. These differences raise important questions regarding how distinct AI-supported learning environments may influence listening development and learner engagement.

Despite its centrality, listening instruction remains underrepresented across many English as a Foreign Language (EFL) settings, especially in Iranian middle schools, where classroom practices tend to prioritize grammar, translation, and standardized test preparation over communicative competence (Ghaed Sharaf et al., 2018; Namaziandost et al., 2019). Exposure to authentic listening materials is often minimal, structured feedback is scarce, and practice opportunities are limited (El Baghdadi et al., 2024). Listening activities frequently serve as isolated evaluation tools rather than cyclical skill-building processes, undermining their pedagogical potential. Moreover, the artificiality of instructional resources —marked by simplified recordings, reduced linguistic variation, and teacher-centered delivery—fails to mirror the complexities of genuine discourse. These limitations highlight a persistent gap between classroom practice and communicative demand, intensifying the need for innovative interventions that enrich auditory input, cultivate sustained engagement, and provide timely, personalized feedback.

In response to these challenges, AI-supported language learning platforms have attracted increasing attention in recent years. Rather than replacing classroom instruction, such technologies are often used to supplement learning by offering structured practice, automated feedback, and interactive learning pathways. Some platforms employ gamified learning structures that incorporate points, levels, and achievement markers to sustain learner engagement, while others emphasize conversational interaction and open-ended dialogue practice. Although these platforms share the goal of supporting language learning, their instructional designs differ in ways that may influence learners' experiences and outcomes in listening development.

Previous studies on AI-assisted language learning have mainly focused on general language proficiency outcomes (Tajik, 2025; Gragera, 2024; Jiang et al., 2023; Jiang & Pajak, 2022), while giving less attention to the relationship between gamified design, personalized feedback, and listening development. Few studies have compared gamified and non-gamified AI tools in adolescent learning contexts, despite the recognized importance of personalization in effective learning (Essafi et al., 2024) and the established role of motivational design in learner engagement (Tajik, 2024). The present mixed-methods study addresses this gap by comparing two different AI-supported learning systems under parallel classroom conditions: a gamified AI-driven personalized learning system (Duolingo) and a non-gamified conversational AI platform (Replika). The study involved 53 Iranian high school students (aged 14–15) who participated in a 12-week instructional program consisting of 24 sessions. Quantitative data were collected through pre- and post-intervention listening tests and learner questionnaires, and qualitative data were obtained through classroom observations and semi-structured interviews. Rather than proposing a new technological model, the study aims to provide a controlled comparison of two

contrasting AI-supported learning environments in order to better understand how differences in instructional design may relate to listening development and learner engagement in an adolescent EFL context.

To provide a conceptual foundation for interpreting these differences, the following section reviews key theoretical perspectives related to listening comprehension, learner self-regulation, scaffolded feedback, and motivational design in technology-supported language learning. These perspectives collectively inform the conceptual orientation of the study and provide the basis for interpreting AI-supported listening instruction through the AELLM and ILED frameworks discussed in the next section.

1.1. Theoretical Framework

1.1.1 Schema Theory

Schema Theory provides an important basis for understanding listening comprehension in EFL contexts because it emphasizes the role of prior knowledge in meaning construction. From this perspective, comprehension is facilitated when learners connect incoming auditory input with existing cognitive schemata, which can reduce processing demands and support interpretation of spoken texts (Xia et al., 2024). In classroom practice, pre-listening activities such as brainstorming, graphic organizers, and KWL strategies can activate relevant background knowledge and help learners anticipate the content of what they are going to hear. Previous research suggests that such activation supports a shift from passive hearing to more active and meaning-focused listening (Xia et al., 2024). In the present study, Schema Theory helps explain why repeated exposure to listening tasks and the activation of prior knowledge may support learners' comprehension, particularly when instructional tasks are organized in a structured way before and during listening practice.

1.1.2 Self-Regulation Theory

Self-Regulation Theory highlights learners' capacity to plan, monitor, and evaluate their own learning in relation to specific goals (Zimmerman, 2002). Self-regulated learners are expected to identify difficulties, apply appropriate strategies, and adjust their efforts on the basis of feedback and performance (Sansone et al., 2019). In digital learning environments, self-regulation can be supported through design features such as goal setting, progress tracking, and repeated practice opportunities (Li & Bonk, 2023). These features are especially relevant in listening instruction, where improvement often depends on sustained practice and learners' ability to reflect on their own progress over time. In the present study, Self-Regulation Theory helps explain how AI-supported environments may encourage learner autonomy, persistence, and strategic engagement with listening tasks, particularly when learners are provided with visible progress indicators and opportunities for repeated practice.

1.1.3 Dynamic Assessment and Sociocultural Theory

Dynamic Assessment, drawing on Vygotsky's Sociocultural Theory, emphasizes the role of interaction, mediation, and feedback in learning (Lantolf, 2000; Lantolf & Poehner, 2014; van Compernolle & Zhang,

2014). Instead of focusing only on learners' independent performance, it considers how learning can develop when appropriate support is provided within the Zone of Proximal Development (ZPD) (Poehner & Lantolf, 2013; Poehner & van Compernolle, 2011; Hidri, 2019). In technology-supported learning environments, this support may take the form of timely feedback, gradual task progression, and guided practice that responds to learner performance (Ma & Zhang, 2024). In the present study, Dynamic Assessment and Sociocultural Theory help explain why feedback and scaffolded support are important in AI-supported listening activities, especially when learners complete tasks that become progressively more challenging over time (van de Pol et al., 2019; Wood, 2021; Swain et al., 2015).

1.1.4 Gamification Theory

Gamification Theory suggests that the use of game-like elements, such as points, levels, rewards, and progress indicators, may enhance learner motivation, persistence, and engagement by making learning tasks more structured and goal-oriented (Shortt et al., 2023). In language learning platforms such as Duolingo, these elements are used to create a sense of progression and achievement, and prior research has linked such features to sustained participation and lower levels of language learning anxiety (García Botero et al., 2018). Gamified design may also reinforce self-regulatory processes by helping learners monitor their performance, set incremental goals, and remain engaged through ongoing feedback. In the present study, Gamification Theory helps explain why differences in motivational design may matter when comparing AI-supported listening environments. More specifically, it provides a basis for examining how points, goals, progression systems, and other motivational features may support learner engagement alongside structured listening practice.

Taken together, these theoretical perspectives suggest that AI-supported listening environments may differ not only in technological format but also in the degree to which they provide prior-knowledge activation, opportunities for repeated and autonomous practice, scaffolded feedback, and motivational support. In this sense, listening development in AI-supported settings can be understood as the result of multiple interrelated instructional dimensions rather than a single technological factor. These dimensions form the conceptual basis of the present study and guide the comparison between the Duolingo group and the Replika group.

Based on this synthesis, the study uses AELLM and ILED not as isolated theoretical labels, but as integrative conceptual models for interpreting AI-supported listening instruction. AELLM draws attention to AI-mediated interaction, feedback, and learner support, whereas ILED emphasizes instructional sequencing, learner engagement, and structured listening development. In this way, both models are grounded in the combined insights of schema activation, self-regulation, scaffolded feedback, and motivational design.

1.1.5 Link to the Research Gap and Aim

Although the theoretical perspectives discussed above explain how prior knowledge activation, learner self-regulation, scaffolded feedback, and motivational design may support language learning, existing research has not sufficiently examined how these dimensions operate across different AI-supported

platforms for EFL listening instruction. In particular, empirical comparisons between gamified and conversational AI environments in adolescent EFL contexts remain limited.

The combined application of Schema Theory, Self-Regulation Theory, Dynamic Assessment within Sociocultural Theory, and Gamification Theory offers a multi layered conceptual lens for examining EFL listening comprehension in AI supported learning environments. Collectively, these perspectives highlight complementary dimensions of listening development, including cognitive preparedness through prior knowledge activation, strategic learner control and autonomy, scaffolded support within the Zone of Proximal Development, and sustained motivational engagement. Although each perspective has been examined independently in prior research (e.g., Gragera, 2024; Jiang et al., 2023; Jiang & Pajak, 2022; Essafi et al., 2024; Fanni & Maharani, 2024), empirical studies that systematically combine these perspectives within authentic classroom contexts—particularly in Iranian secondary education—remain limited.

More specifically, few studies have examined how these theoretical dimensions are reflected in the design features of different AI-supported language learning platforms. As a result, the ways in which schema activation, self-regulated learning behaviors, scaffolded feedback, and motivational gamification interact within such systems are not yet fully understood. This limitation creates a gap between theoretical explanations of listening development and the practical design of AI-supported learning environments.

To address this gap, the present study interprets AI-supported listening instruction through two complementary conceptual frameworks: the AI Enhanced Language Learning Matrix (AELLM) and the Intelligent Learning Environment Design (ILED). These frameworks function as organizing structures that map cognitive, regulatory, sociocultural, and motivational processes onto the design features of AI supported listening instruction.

Building on the empirical patterns observed in the present study, the analysis is further interpreted through Tajik's Scaffolded Motivator Model (T-SMM), which is introduced as a context-specific explanatory framework describing how feedback and motivational scaffolding may support learner engagement and listening development over time.

Conducted with adolescent EFL learners in Iranian middle schools, the present research examines how these theoretically grounded mechanisms can be organized within AI supported learning environments to deliver structured, engaging, and developmentally appropriate listening instruction. By linking theoretical synthesis with empirical comparison, the study aims to clarify how differences in AI-supported instructional design may relate to listening development and learner engagement in adolescent EFL contexts.

In addition to informing the instructional design, the theoretical perspectives also guided the development of the research instruments used in this study. Constructs derived from Self-Regulation Theory and Gamification Theory informed the design of the learner perception questionnaire by targeting

dimensions such as goal setting, progress monitoring, motivational persistence, and perceived usefulness of feedback. Schema Theory and Sociocultural principles underlying Dynamic Assessment shaped the semi structured interview protocol, which elicited learners' reflections on pre listening preparation, adaptive support, and scaffolded task progression. Furthermore, the classroom observation checklist was aligned with sociocultural and gamification constructs by capturing behavioral indicators of engagement, structured progression, learner–system interaction, and adaptive response to feedback. Through this alignment, the theoretical framework was operationalized across instructional design, data collection instruments, and analytical procedures, ensuring coherence between conceptual foundations and the empirical investigation.

2. Literature Review

Building on the integrated theoretical perspectives from Section 1.1.5, this chapter synthesizes empirical findings and conceptual discussions on gamified and non-gamified AI-driven language learning platforms in the context of EFL listening comprehension. The review adopts a thematic structure encompassing Duolingo's technical and pedagogical features, the motivational and skill-development affordances of gamification, evidence on Duolingo's role as an instructional intervention, and empirical insights into the intersection of gamification, intelligent platforms, and listening skill development. This synthesis situates the present study within ongoing scholarly debates while identifying persistent theoretical and practical gaps that underlie its design.

2.1 Duolingo and Its Features

The rapid diffusion of mobile-assisted and AI-supported language learning has redefined accessibility, allowing learners to continue their studies independently of traditional classroom constraints. Among language learning has redefined accessibility, allowing learners to continue their studies independently of traditional classroom constraints. Among such technologies, **Duolingo** represents a widely recognized benchmark, supporting more than 23 languages and engaging over 200 million users globally (Jašková, 2014). Its gamified interface delivers adaptive tasks spanning listening, speaking, reading, and writing domains (Inayah et al., 2020). Through placement tests calibrated to individual proficiency (Nushi & Eqbali, 2017), Duolingo constructs personalized entry points that align with learners' current linguistic competence.

Features such as streaks, crown levels, and leaderboards function not merely as decorative mechanics but as structured motivational cues fostering sustained participation and progress monitoring (Jiang et al., 2024; García Botero et al., 2018; Qub'a et al., 2024; Szabó & Kopinska, 2023). These systems encourage consistent engagement through feedback-driven goal cycles, effectively operationalizing principles from Self-Regulation Theory. At a more technical level, Duolingo's analytics track fine-grained sub-skills such as phoneme recognition or prosodic awareness (Tuong & Dan, 2024), allowing adaptive recalibration of content difficulty. Consequently, it provides a data-driven model of individualized scaffolding, linking AI-based adaptivity with pedagogical accountability. In short, Duolingo illustrates how

gamified architecture and algorithmic adaptivity converge to form a controlled environment for examining the intersection of engagement, feedback, and skill development.

2.2 The Role of Gamification in Enhancing EFL Listening Skills

Gamification extends beyond the superficial inclusion of games; it embodies the systematic application of game-design logic—points, rewards, challenges, and progression—to learning processes. In EFL pedagogy, this logic has consistently shown potential to stimulate motivation and persistence (Bennani et al., 2022; Hsu, 2024). Yet the significance of gamification lies not in isolated features but in how it cultivates self-directed habits through structured cycles of goal-setting and feedback (García-Botero et al., 2018). When embedded in adaptive systems, game-based elements trigger both extrinsic momentum (through competition or badges) and intrinsic satisfaction (through perceived achievement and autonomy) (Shortt et al., 2023).

These mechanisms are particularly relevant to listening comprehension, a skill often constrained by transient auditory input and cognitive load. Gamified drills—such as timed listening challenges or repetition streaks—promote attentional control and resilience for processing connected speech and prosodic patterns (Bakhtiar et al., 2024). By integrating such cycles, learners regulate their progress, set incremental listening goals, and evaluate performance, mirroring self-regulation in action (Zimmerman, 2002).

Furthermore, studies suggest that in multicultural settings, gamified collaboration can reduce anxiety and heighten willingness to communicate (García-Botero et al., 2019), consistent with sociocultural models of engagement (Swain et al., 2015). Therefore, gamification contributes not only to behavioral persistence but also to social-emotional readiness—two preconditions for meaningful listening development. Overall, its educational promise arises from the synergy between motivational architecture and cognitive adaptability, a linkage central to this study's exploration of gamified learning systems.

2.3 Duolingo's Effectiveness as a Language Learning Tool

Despite Duolingo's global reach, empirical investigation into its **skill-specific impacts**, particularly in listening, remains limited. Most prior research has concentrated on general proficiency outcomes, rarely dissecting the cognitive or motivational mechanisms responsible for them (García-Botero et al., 2020; Zhang & Hasim, 2023). Yet several of its design features—adaptive sequencing, immediate corrective feedback, and the repetition of authentic audio input—have clear theoretical grounding. Collectively, these align with **Schema Theory** and **Dynamic Assessment** principles: learners activate prior schemata, receive scaffolded correction, and gradually internalize patterns of decoding and comprehension (Ghasemi et al., 2024; Poehner & van Compernolle, 2011).

The **microlearning format** consolidates knowledge through distributed practice, echoing cognitive findings on spaced repetition and retention. Similarly, the inclusion of rich contextual cues situates comprehension within meaningful discourse contexts (Apio, 2022). However, critical gaps persist.

Studies seldom scrutinize how **gamified feedback loops** mediate listening gains or how adaptation interacts with learner diversity—age, proficiency, or cultural background. Without such specification, claims about Duolingo's listening efficacy remain associative rather than explanatory.

Addressing this limitation requires comparative designs that distinguish whether observed improvements derive from **gamification dynamics**, **adaptive algorithms**, or their interaction. The present study responds directly to that call by examining Duolingo alongside a **non-gamified conversational AI platform**, thereby clarifying the pedagogical weight of gamified design within AI-supported listening development.

2.4 Empirical Studies on Duolingo, Gamification, and Listening Skills

Empirical research on gamified intelligent learning platforms has consistently reported heightened learner engagement and, in certain contexts, measurable gains in language skills. However, these outcomes appear to be strongly mediated by contextual, pedagogical, and learner-related variables rather than by technological affordances alone. Lee (2019), for instance, emphasizes that the effectiveness of digital learning tools is shaped by socio-political and instructional conditions, cautioning against uncritical generalization of positive findings across educational contexts. Similarly, Tai and Chen (2020) demonstrate that AI-supported assistants can enhance learners' confidence and perceived competence—affective factors that may indirectly support listening development but do not, in isolation, account for skill-specific gains.

Studies focusing specifically on Duolingo report improvements in discrete listening sub-skills, including phoneme discrimination, auditory recognition, and basic decoding processes. Nevertheless, much of this research is constrained by short intervention periods, limited longitudinal tracking, or the absence of rigorous control-group designs, thereby restricting causal interpretation (Jiang et al., 2024). Consequently, while Duolingo's effectiveness is frequently reported, the pedagogical mechanisms through which gamified design features and adaptive instructional feedback contribute to sustained listening comprehension development remain under-theorized and empirically fragmented.

Parallel strands of research adopting Dynamic Assessment frameworks highlight the instructional value of responsive, scaffolded feedback in listening pedagogy (Huynh & Iida, 2016; Dehghanzadeh et al., 2021; Ma & Zhang, 2024). These studies suggest that mediation aligned with learners' developmental needs can facilitate listening growth more effectively than static assessment practices. However, such approaches are rarely examined in conjunction with gamified intelligent learning systems, leaving unresolved questions regarding how scaffolded instructional feedback and motivational design interact within AI-driven platforms.

Crucially, comprehensive empirical comparisons between **gamified** and **non-gamified** intelligent learning systems—particularly among adolescent EFL learners—remain scarce. This gap is especially pronounced in the Iranian educational context, where listening instruction is shaped by curriculum constraints,

examination-oriented practices, and limited exposure to authentic auditory input. Existing studies tend to isolate either gamification effects or AI-mediated interaction, rather than systematically contrasting these instructional architectures under controlled classroom conditions.

In summary, although gamified AI-driven systems such as Duolingo show promise for enhancing listening comprehension through adaptive feedback, motivational scaffolding, and targeted practice, the empirical evidence base remains context-bound and methodologically uneven. Few studies explicitly examine the interplay among gamification mechanisms, self-regulatory learning behaviors, and quantifiable listening outcomes in adolescent populations. This absence of robust, context-sensitive comparative evidence—particularly in Iranian secondary education—constitutes the empirical gap addressed by the present study. By employing a controlled, mixed-methods design with Iranian adolescent EFL learners, this research advances understanding of how gamified and non-gamified intelligent systems may differentially support both the cognitive and affective dimensions of listening development.

Therefore, our true experimental study of 53 Iranian secondary school students bridges this gap by examining how Duolingo's gamified intelligent learning system enhances listening comprehension through adaptive algorithms and engagement strategies. The experimental group demonstrated significant improvements (27.8% increase) compared to traditional methods (8.3% increase). Building on these theoretical foundations and addressing identified research gaps, this study hypothesizes that integrating AI-driven emotional intelligence in language learning platforms is positively associated with improved speaking performance (H1).

Research Questions

1. Are there significant differences between the effects of a gamified AI-driven personalized learning system (Duolingo) and a non-gamified conversational AI platform (Replika) on middle school EFL students' listening comprehension proficiency and engagement levels?
2. How do middle school students perceive Duolingo's personalized learning paths as an effective means of enhancing their listening comprehension and engagement in EFL learning?
3. Do the results obtained from classroom observation checklists in the Duolingo experimental group corroborate the findings derived from learner interviews and perception questionnaires?

The following null hypothesis was tested statistically to address the first research question of the study:

H0: No significant differences exist between the effects of the gamified AI-driven personalized learning system (Duolingo) and the non-gamified conversational AI platform (Replika) on middle school EFL students' listening comprehension proficiency and engagement levels.

3. Methodology

This study employed a mixed-methods approach using a concurrent triangulation design to examine the effectiveness of two AI-supported learning platforms, Duolingo and Replika, in improving listening comprehension among Iranian middle school students learning English as a foreign language (EFL). Quantitative and qualitative data were collected during the same period and integrated during interpretation to provide a comprehensive understanding of learning outcomes and student experiences.

Participants consisted of 53 students aged 14–15 years from public middle schools in Tehran Province. Students were randomly assigned to either the Duolingo group (n = 27) or the Replika group (n = 26). Both groups participated in a 12-week instructional intervention comprising 24 sessions delivered through their assigned AI-supported platform. The instructional activities focused on listening comprehension tasks aligned with the national English curriculum.

3.1 Instruments and Data Collection

3.1.1 Listening Comprehension Assessment

To examine listening development and ensure baseline comparability between groups, pre-test and post-test listening comprehension assessments were administered. The tests were designed for Iranian eighth- and ninth-grade learners and aligned with the *Prospect 2* and *Prospect 3* English language curricula. The assessment consisted of five sections: Understanding Main Conversations (5 marks), Specific Information Detection (4 marks), Classroom Instructions Comprehension (3 marks), True/False Recognition (4 marks), and Dialogue Completion (4 marks), producing a total score out of 20.

For statistical analysis, a composite listening score was calculated based on learners' performance across the assessment components to provide a stable indicator of listening comprehension. Content validity was established through expert evaluation by three specialists in English language teaching and language assessment, who confirmed the alignment of the test items with curriculum-based listening objectives. Reliability was examined through a pilot study with 30 students with similar characteristics to the study sample. The analysis yielded a Cronbach's alpha coefficient of 0.87, indicating acceptable internal consistency.

3.1.2 Perception Questionnaire

A researcher-developed questionnaire was used to examine students' perceptions of AI-supported learning. The instrument included 18 items across five dimensions: learning engagement, system usability, perceived effectiveness, learning progress, and motivational aspects. Responses were recorded using a five-point Likert scale ranging from strongly agree to strongly disagree.

Construct validity was examined through exploratory and confirmatory factor analyses. The results indicated an acceptable factor structure, and the internal consistency of the questionnaire was satisfactory (Cronbach's alpha = .89). The questionnaire also collected information on students'

frequency of system use, completion of assigned activities, and voluntary engagement in additional listening tasks.

3.1.3 Semi-Structured Interviews

To complement the quantitative findings, semi-structured interviews were conducted after the intervention. A purposive sample of 20 students was invited to participate, and 15 interviews yielded complete and analyzable data. The interview protocol consisted of eight questions designed to explore learners' experiences with AI-supported learning systems, including perceived learning benefits, motivation, engagement, and self-confidence in listening activities.

Each interview lasted between 25 and 35 minutes and was audio-recorded with participants' consent. The recordings were transcribed for analysis. To enhance credibility, participants were given the opportunity to review summaries of their responses (member checking), and the coding process was discussed with an experienced EFL researcher (peer debriefing).

3.1.4 Observation Checklists

Classroom observations were conducted following the intervention to document learner interaction patterns and engagement during AI-supported learning activities. Observations were guided by a structured checklist developed for this study (Appendix C), focusing on indicators such as learner participation, interaction with the digital platform, and responsiveness to listening tasks.

Two trained observers independently recorded classroom behaviors across twenty instructional sessions. Observations were coded using a binary system accompanied by descriptive field notes to capture contextual information. Inter-rater reliability was calculated prior to the main analysis using Cohen's kappa, resulting in a coefficient of 0.88, which indicates a high level of agreement. Any differences in coding were discussed and resolved through consensus.

3.2 Data Analysis

Quantitative data were analyzed using SPSS. Paired-samples t-tests were conducted to examine within-group changes between pre-test and post-test listening scores. Independent-samples t-tests were used to compare post-intervention listening performance between the Duolingo and Replika groups. In addition, one-way ANOVA was applied where appropriate to examine differences across relevant variables.

Qualitative data from interviews and observation notes were analyzed using thematic analysis following the procedures described by Braun and Clarke (2006). The analysis involved several stages. First, all transcripts and observation notes were read repeatedly to achieve familiarity with the data. Second, initial codes were generated using a hybrid coding approach. Deductive codes were informed by the study's conceptual framework and questionnaire dimensions, while additional codes were identified inductively from the participants' responses and observation records.

Coding was conducted manually to allow close examination of the data. After initial coding, related codes were grouped into broader categories, and preliminary themes were identified. These themes were then reviewed and refined through comparison with the original data to ensure internal coherence and conceptual clarity. Final themes were defined and organized to represent recurring patterns in students' experiences with AI-supported learning.

To strengthen the trustworthiness of the qualitative analysis, several strategies were employed. These included methodological triangulation across interview data, observation records, and questionnaire responses, member checking with selected participants, and peer review of the coding structure. These procedures helped ensure that the qualitative findings were grounded in the data and supported by multiple sources of evidence.

The integration of quantitative and qualitative findings allowed the study to provide both statistical evidence of learning outcomes and a contextual understanding of how students experienced the use of AI-supported language learning platforms.

3.1. Participants

This study was conducted in Varamin, a city located southeast of Tehran, Iran, and involved middle school students aged 14–15 years enrolled in grades 8 and 9. A multi-stage sampling procedure was employed to obtain a representative sample. Initially, eligible lower secondary schools were identified from an official registry, and six schools representing both male and female student populations were selected through stratified randomization to ensure balanced demographic representation.

All prospective participants (N = 195) completed a standardized language proficiency test to assess baseline language skills. Systematic screening criteria were then applied, including intermediate-level proficiency, regular attendance, access to mobile learning devices, and parental consent. Students who did not meet these criteria were excluded during the screening stage. Therefore, the reduction from 195 students to the final sample of 53 was due to the application of the inclusion criteria, not participant attrition. Based on these criteria, 53 students met the inclusion requirements and were selected as the final sample. They were randomly assigned to the Duolingo group (n = 27) or the Replika group (n = 26). Identical demographic and academic inclusion criteria were applied across both groups to control for potential confounding variables. No participant attrition occurred during the intervention period.

The 12-week instructional intervention comprised 24 classroom sessions. The Duolingo group used a gamified digital learning platform incorporating features such as leaderboards, point-based progress tracking, and interactive challenges to support listening comprehension through structured and task-based activities. The comparison group used the Replika AI application, a conversational agent designed to simulate natural dialogue and provide language practice through personalized interaction. Both groups received equivalent instructional content and duration, ensuring that observed differences could be attributed to platform design rather than instructional exposure. This design enabled a focused

comparison of gamified and conversational AI-based learning tools in relation to EFL listening development while minimizing demographic and instructional confounds.

3.2. Data Collection Instruments

Before detailing the data collection instruments, it is necessary to clarify how the key conceptual constructs of the study were operationalized. In this research, human-centered artificial intelligence is not treated as a claim about internal algorithmic architectures or proprietary system intelligence. Rather, it is operationalized as a pedagogical orientation manifested through learner-facing features such as system usability, learner agency, progress transparency, self-paced learning, and motivational feedback. These dimensions were examined through learners' questionnaire responses, semi-structured interviews, and systematic classroom observations.

Similarly, emotion-adaptive AI is operationalized at the experiential and instructional level rather than at the computational level. No emotional signal detection, affective computing, or physiological data processing was accessed or manipulated during the intervention. Instead, emotional adaptivity refers to learners' perceived emotional and motivational responsiveness of the learning system, inferred from observable indicators such as encouraging feedback, persistence behaviors (e.g., sustained use and task completion), adaptive task sequencing, and learners' reflective reports of emotional support and motivation. These indicators were systematically documented through observation checklists and post-intervention interviews, allowing emotional adaptivity to be examined as a pedagogical and perceptual construct rather than an algorithmic mechanism.

This study adopted a mixed-methods research design to evaluate the comparative effectiveness of **two AI-supported learning systems—Duolingo and Replika—** in enhancing EFL listening comprehension. Multiple instruments were employed to ensure comprehensive evaluation and methodological triangulation.

The primary quantitative measure was a pre-- and post-intervention listening comprehension assessment specifically developed for Iranian 8th- and 9th-grade students, based on *Prospect 2* and *Prospect 3* textbooks. The test consisted of five sections: (1) Comprehension of main conversations (5 marks), (2) Recognition of specific information (4 marks), (3) Comprehension of classroom instructions (3 marks), (4) True/False recognition (4 marks), and (5) Dialogue completion (4 marks), yielding a maximum score of 20 marks.

A rigorous validation process confirmed the instrument's psychometric soundness. Content validity was established by a panel of nine experts—six university professors, two English language supervisors, and one experienced trainer—resulting in a Content Validity Index (CVI) = 0.90 and a Content Validity Ratio (CVR) = 0.88. Construct validity was supported by factor analysis (KMO = 0.83; Bartlett's test, $p < .001$), which identified five components corresponding to the test sections. Convergent validity was evidenced by a strong correlation with standardized listening tests ($r = 0.85$).

Reliability indices were equally robust: Cronbach's alpha ranged from 0.81 to 0.85 across sections (0.87 overall), inter-rater reliability achieved a Cohen's Kappa = 0.89 (Pearson *r* = 0.92), and test-retest reliability over two weeks with 40 students yielded *r* = 0.86. Item analysis demonstrated appropriate difficulty indices (0.35–0.75), discrimination indices (0.38–0.65), and point-biserial correlations (0.42–0.68). Expert feedback informed refinements related to timing (25 minutes total), instruction clarity, content alignment, audio quality, and repetition frequency (each section played twice, 30 seconds apart). The test content differentiated grade-specific complexity, ranging from everyday conversations (Grade 8) to advanced topics and complex structures (Grade 9), ensuring cultural appropriateness and precise targeting of Iranian EFL learners' abilities.

The second quantitative instrument was an 18-item researcher-developed questionnaire (Appendix A), measuring learning engagement, system usability, perceived effectiveness, learning progress, and motivational aspects on a five-point Likert scale (*Strongly Agree* to *Strongly Disagree*). Validity and reliability were confirmed through exploratory and confirmatory factor analyses (α = .89). The questionnaire captured data on learners' engagement with the AI learning systems, including frequency of application use, homework completion, and voluntary participation in listening activities.

Two qualitative instruments provided complementary insights. First, semi-structured post-intervention interviews were conducted with a purposive subsample of 15 participants (Appendix B) to explore motivation, self-efficacy, and perceived benefits of personalized learning paths. These participants were selected based on their availability, level of engagement during the intervention, and willingness to articulate reflective learning experiences. Interview sessions (25–35 minutes) were audio-recorded, transcribed verbatim, and validated through member checking and peer debriefing. Although an initial pool of 20 learners was invited to participate in the interview phase, complete and analyzable interview data were obtained from 15 participants. Accordingly, only these 15 interviews were included in the qualitative analysis. Second, a structured observation checklist (Appendix C), validated by experts and pilot-tested (inter-rater agreement = 0.88), tracked learner–system interactions, environment dynamics, and engagement indicators across 20 sessions, with binary coding supplemented by qualitative notes.

This integrated multi-instrument approach enabled both quantitative outcome measurement and rich qualitative exploration of the learning process, supporting a nuanced analysis of the cognitive and affective dimensions of technology-enhanced listening comprehension among adolescent EFL learners.

3.3. Data Collection Procedure

To select participants, the researchers administered a standardized version of the Preliminary English Test (PET) to 195 middle school students aged 14–15 from Varamin County, Iran, to determine their baseline proficiency. Following a comprehensive assessment, 53 learners with comparable intermediate language skills were included in the study. These participants were then randomly assigned to **the Duolingo group (n = 27)** and **the Replika group (n = 26)**. This group allocation ensured comparable baseline proficiency across the two AI-supported instructional conditions, thereby minimizing the influence of pre-existing differences on the study outcomes.

Data collection was carried out in four distinct phases. **In the first phase**, midway through the second academic term, both groups underwent a listening test to assess their pre-experiment comprehension levels. The experimental group then engaged with the Duolingo platform through regular task performance, completing tasks both inside and outside the classroom. In a structured 12-session programme using Duolingo, middle school students in Iran followed an engaging and gamified learning experience, with each session designed to build on the previous one while introducing new challenges to reinforce essential listening skills. For example, in the first session, students worked on a listening exercise that involved associating simple words with images, such as matching the sound of the word "apple" with a picture of an apple. This activity helped them to develop their auditory recognition intuitively. In the second session, the focus shifted to short sentences, where students listened to sentences such as "The cat is on the mat" and selected the correct written form from multiple-choice options, successfully linking listening comprehension with sentence structure.

In subsequent sessions, more dynamic activities were introduced, such as story-based listening tasks (session 3), where students listened to short stories and answered questions such as "Where did the boy go?" or "What did he buy?", which promoted deeper retention and listening focus. The fourth session introduced real-life conversations where students listened to dialogues (e.g., ordering food in a restaurant) and practiced repeating sentences to improve pronunciation and conversational skills. As the programme progressed, Sessions 5 and 6 introduced greater complexity through different accents and speeds, asking students to complete tasks such as filling in blanks in dialogues spoken in British or American accents. Session 7 utilized Duolingo's review features, allowing students to revisit previously challenging phrases or exercises to strengthen weak areas through highly personalized practice. Finally, the entire session included a gamified listening assessment where students answered questions based on longer dialogues or stories, allowing them to measure their progress and celebrate their achievements.

Throughout this programme, Duolingo's gamification elements - such as earning XP points for completing tasks, tracking progress on leaderboards, and earning badges for milestones - kept students consistently motivated and engaged, while its streak feature encouraged regular practice to solidify improvements. In contrast, the comparison group used Replika, an AI-driven tool focused on natural and dynamic conversation. Unlike Duolingo, Replika does not include gamified elements such as points or badges, allowing students to improve their listening and conversational skills in a more organic, non-competitive environment. This setup provided an opportunity to compare the engaging, game-like structure of Duolingo with the conversational depth and adaptability of a non-gamified AI system.

To ensure procedural comparability, both the experimental and comparison groups participated in the intervention over the same instructional duration (12 weeks, 24 sessions) and under equivalent classroom scheduling conditions. However, it is important to acknowledge that the **nature and structure of learning activities differed inherently between the two AI platforms**, reflecting their distinct pedagogical architectures. While the Duolingo environment employed highly structured, task-based listening drills with adaptive progression and gamified feedback, the Replika condition emphasized open-

ended conversational interaction without predefined task sequences or gamified reinforcement. These differences were not treated as extraneous variables but constituted an **intentional component of the intervention design**, allowing for a comparative examination of two contrasting AI-supported instructional approaches. Accordingly, the findings should be interpreted as reflecting differences between **integrated instructional configurations** rather than the isolated effect of any single design feature (e.g., gamification or adaptivity alone).

In the second phase, the experimental group was administered an 18-item researcher-developed perception questionnaire designed to measure five dimensions: learning engagement, system usability, perceived effectiveness, learning progress, and motivational aspects. This instrument used a five-point Likert scale and demonstrated high reliability (α = .89) through both exploratory and confirmatory factor analyses.

In the third phase, semi-structured interviews lasting 25–35 minutes were conducted with 15 participants immediately after the treatment period. These interviews provided rich, detailed insights into participants' experiences with Duolingo's gamified features, focusing on learning motivation, self-efficacy, and perceived benefits of personalized learning paths. All interviews were digitally recorded and transcribed verbatim for analysis.

The fourth phase used a researcher-developed classroom observation checklist to assess learner-system interaction patterns, learning environment dynamics, and engagement indicators. Observations were conducted over twenty sessions during the twelve-week intervention period, using a binary coding system supplemented by qualitative notes. Following the treatment period, a post-test was administered to assess the impact of the intervention. Independent-samples t-tests were conducted in SPSS to compare the post-intervention listening comprehension performance between the Duolingo and Replika groups. The study adhered to ethical protocols, including informed consent, confidentiality, and participants' right to withdraw. At the same time, the multiple data collection instruments facilitated methodological triangulation for a comprehensive analysis of the effectiveness of AI-integrated language teaching.

3.4 Qualitative Data Analysis and Trustworthiness

Qualitative data obtained from semi-structured interviews and classroom observations were analyzed using thematic analysis, following the systematic procedures outlined by Braun and Clarke (2006). A hybrid coding approach was adopted, combining deductive and inductive strategies. Initially, a deductive coding framework was informed by the study’s theoretical foundations (e.g., self-regulation, scaffolding, gamification, and comprehensible input) and the constructs measured in the perception questionnaire. Subsequently, inductive coding was applied to allow additional patterns and themes to emerge directly from participants’ responses and observational records.

The analysis proceeded through six iterative phases: (1) familiarization with the data through repeated reading of interview transcripts and observation notes; (2) generation of initial codes capturing salient

features related to listening engagement, strategy use, and interaction with the AI platform; (3) clustering related codes into preliminary themes; (4) reviewing themes across data sources to ensure internal coherence and distinctiveness; (5) defining and refining final themes; and (6) reporting themes with representative excerpts and frequency-based observational indicators. Coding was conducted manually to maintain close engagement with the data.

To enhance the trustworthiness of the qualitative findings, multiple strategies were employed. Methodological triangulation was achieved by comparing themes across interviews, observation checklists, and perception questionnaire results. Member checking was conducted by sharing summarized interpretations with selected interview participants to confirm the accuracy of meaning. Peer debriefing involved consultation with an experienced EFL researcher who reviewed the coding scheme and thematic structure for clarity and consistency.

For classroom observations, two trained observers independently coded behavioral indicators using a structured checklist. Inter-coder agreement was calculated prior to full analysis, yielding a reliability coefficient of 0.88, indicating high consistency. Any discrepancies were resolved through discussion and consensus. Collectively, these procedures ensured analytic rigor, transparency, and credibility, enabling the qualitative findings to complement and substantiate the quantitative results within the mixed-methods design.

4. Results

4.1. Assessment of Initial Listening Comprehension: A Pre-intervention Analysis

Before the implementation of the experimental intervention, all participants underwent a comprehensive listening comprehension pre-test designed in alignment with the Prospect 2 and Prospect 3 curricula for Iranian lower secondary education. The instrument encompassed five discrete subskills—Main Conversation Comprehension, Specific Information Identification, Classroom Instruction Understanding, True/False Statement Analysis, and Dialogue Completion Tasks—yielding an aggregated score representing overall listening proficiency.

Independent samples t-test analyses conducted on the pre-test outcomes (Table 1) demonstrated that there were no statistically significant differences between the Duolingo group (n = 27) and the Replika group (n = 26) across any of the five subskills or in total listening scores (all p-values > .05). Specifically, mean scores for Main Conversation Comprehension were 4.98 (SD = 0.96) for the experimental group and 5.93 (SD = 1.10) for the comparison group ($t(51) = 1.889$, $p = .066$). For Specific Information Identification, mean scores were 4.92 (SD = 1.01) and 5.34 (SD = 0.68), respectively ($t(51) = 1.116$, $p = .268$). Scores for Classroom Instruction Understanding were identical in mean (4.89) between groups, evidencing complete parity at baseline ($t(51) = 0.973$, $p = .335$). True/False Statement Analysis yielded means of 4.28 (SD = 1.30) and 4.91 (SD = 1.02) ($t(51) = 1.565$, $p = .124$), whereas Dialogue Completion

Tasks averaged 4.42 (SD = 1.21) and 4.88 (SD = 1.03) (t(51) = 1.793, p = .078). Regarding the overall performance, total pre-test scores averaged 26.02 (SD = 4.46) for the Duolingo group and 24.12 (SD = 3.46) for the Replika group (t(51) = 1.899, p = .064).

Although some directional differences were observable across several listening subskills at the pre-test stage, none of these differences reached statistical significance (p > .05). Therefore, the two groups were considered sufficiently comparable prior to the intervention, allowing subsequent analyses to focus on post-test performance differences.

These non-significant findings support the statistical comparability of the two cohorts at the outset of the study. This comparability enhances the confidence with which subsequent changes in performance may be considered in relation to the respective interventions—Duolingo's gamified, rewards-based environment or Replika's conversation-oriented, context-rich dialogue system—rather than solely attributable to pre-existing disparities in listening comprehension ability.

4.2. The Results of the First Research Question

As shown in Table 2, both groups demonstrated improvements in post-test listening comprehension scores; however, the Duolingo-based experimental group consistently outperformed the Replika comparison group across all measured subskills. It should be noted that the reported listening scores represent **composite listening performance derived from multiple listening tasks administered over the instructional period**, rather than raw scores from a single 20-point test.

For **Main Conversation Comprehension**, the experimental group achieved a higher mean score (M = 6.55, SD = 0.508) than the comparison group (M = 5.72, SD = 0.994), and this difference was statistically significant, $t(51) = 4.746$, $p < .001$, **Cohen's d = 1.06**, indicating a **large effect size**.

In **Specific Information Identification**, Duolingo participants obtained a mean score of 6.71 (SD = 0.739), compared with 5.42 (SD = 0.867) for Replika participants. This difference was statistically significant, $t(51) = 5.661$, $p < .001$, with a **large effect size** (**d = 1.60**).

Similarly, **Classroom Instruction Understanding** scores were significantly higher in the experimental group (M = 6.41, SD = 0.565) than in the comparison group (M = 5.35, SD = 0.898), $t(51) = 4.226$, $p < .001$, **d = 1.42**, again reflecting a **substantial instructional effect**.

Although the performance gap was smaller for **True/False Statement Analysis**, the experimental group still outperformed the comparison group (M = 5.91, SD = 0.679 vs. M = 5.51, SD = 1.152), and this difference reached statistical significance, $t(51) = 2.297$, $p = .026$, with a **small-to-moderate effect size** (**d = 0.43**).

The **largest difference** was observed in **Dialogue Completion Tasks**, where the experimental group achieved a mean of 6.21 (SD = 0.478), compared with 5.08 (SD = 0.752) for the comparison group, $t(51) = 5.873$, $p < .001$, **d = 1.80**, indicating a **very large effect**.

Regarding **total post-test listening scores**, the Duolingo group obtained a substantially higher composite score (M = 31.32, SD = 1.751) than the Replika group (M = 26.42, SD = 3.620), and this difference was statistically significant, $t(51) = 6.085$, $p < .001$, with a **large overall effect size** (**d = 1.73**). Collectively, these findings demonstrate that the gamified, AI-driven Duolingo platform produced **robust and practically meaningful gains** in EFL listening comprehension compared to the primarily conversational Replika platform.

To further examine within-group performance patterns, a one-way ANOVA was conducted for the experimental group (Table 3). The results revealed no statistically significant differences among the five listening subskills, $F(4, 125) = 2.317$, $p = .078$, suggesting that Duolingo's instructional impact was **evenly distributed across skill areas**. This balanced improvement supports the interpretation that Duolingo's adaptive and competitive features—such as leaderboards, goal tracking, and immediate feedback—foster **holistic listening comprehension development**, rather than selectively enhancing particular task types. In contrast, Replika's focus on conversational interaction, while beneficial for engagement, lacks the structured gamification mechanisms that appear to drive sustained proficiency gains.

Table 1

Means, Standard Deviation, and T-test. The Effects of the Experimental and Comparison Groups on (Pre) Student Performance on the Listening Comprehension Test

Main Conversation Comprehension	**GROUP**	**N**	**Mean**	**SD**	**T**	**df**	**Sig**
	Experimental	**27**	**4.98**	**.956**	**1.889**	**51**	**.066**
	Comparison	26	5.93	1.102			
Specific Information Identification	Experimental	27	4.92	1.009	1.116	51	.268
	Comparison	26	5.34	.675			
Classroom Instruction Understanding	Experimental	27	4.89	1.087	.973	51	.335
	Comparison	26	4.89	.891			
True/False Statement Analysis	Experimental	27	4.28	1.302	1.565	51	.124
	Comparison	26	4.91	1.024			
Dialogue Completion Tasks	Experimental	27	4.42	1.208	1.793	51	.078
	Comparison	26	4.88	1.029			
Total Scores of Listening Pre-test	Experimental	27	26.02	4.456	1.899	51	.064
	Comparison	26	24.12	3.458			

Table 2

Means, Standard Deviations, and T-test results from the Student's Post-Listening Comprehension Test for the Experimental and Comparison Groups

Main Conversation Comprehension	GROUP	N	Mean	SD	T	df	Sig
	Experimental	**27**	**6.55**	**0.508**	**4.746**	**51**	**0.000**
	Comparison	26	5.72	0.994			
Specific Information Identification	Experimental	27	6.71	0.739	5.661	51	0.000
	Comparison	26	5.42	0.867			
Classroom Instruction Understanding	Experimental	27	6.41	0.565	4.226	51	0.000
	Comparison	26	5.35	0.898			
True/False Statement Analysis	Experimental	27	5.91	0.679	2.297	51	0.026
	Comparison	26	5.51	1.152			
Dialogue Completion Tasks	Experimental	27	6.21	0.478	5.873	51	0.000
	Comparison	26	5.08	0.752			
Total Scores of Listening Post-test	Experimental	27	31.32	1.751	6.085	51	0.000
	Comparison	26	26.42	3.620			

Table 3

One-Way ANOVA Results of the Experimental Group Students' Listening Comprehension Aspects

Between Groups	Sum of Squares	df	Mean Square	F	Sig.
	8.214	**4**	**2.054**	**2.317**	**.078**
Within Groups	110.762	125	0.886		
Total	118.976	129			

4.2. Results of the Second Research Question

The second research question explored middle school students' perceptions and experiences regarding Duolingo's personalized learning paths as a pedagogical intervention to enhance EFL listening comprehension. Employing a mixed-methods design, data were collected via both structured questionnaires and semi-structured interviews, ensuring a comprehensive understanding of learner engagement and response to the platform's features.

The questionnaire targeted multiple interconnected dimensions: (1) neural-reward-based gamification elements, (2) AI-driven adaptive scaffolding, (3) multi-modal presentation and repetition strategies, (4) social learning dynamics, (5) human–computer interaction features, and (6) situated learning

algorithms. This multidimensional approach enabled a nuanced examination of the motivational, cognitive, and interactional mechanisms underlying learner engagement.

The analysis is presented in two sequential phases. First, quantitative findings from the questionnaire are reported, highlighting measurable trends in learners' perceptions across the aforementioned dimensions. This is followed by qualitative insights derived from interview data, offering depth and contextualization to the numerical patterns. Together, these complementary perspectives provide a robust account of how students interacted with—and benefited from—Duolingo's innovative, personalized learning pathways in developing their listening comprehension proficiency.

4.2.1 Results of the questionnaire

Analysis of questionnaire responses provided detailed insights into students' perceptions of Duolingo's personalized learning paths for enhancing EFL listening comprehension.

In the dimension of **reward-based gamification**, Item 10 (M = 4.32, SD = 0.892) indicated high engagement with features such as streak tracking and achievement badges, which participants associated with sustained motivation. Regarding **adaptive scaffolding**, Items 6 and 12 (M = 4.21, SD = 0.934) reflected positive learner perceptions of dynamic difficulty adjustment and individualized progress monitoring within the platform.

The dimension related to **instructional strategies inspired by language learning theories**, measured in Item 8 (M = 4.19, SD = 0.912), was associated with perceived improvements in vocabulary retention through multimodal listening activities. The **social learning** dimension, captured by Item 13 (M = 4.15, SD = 1.023), highlighted the motivational influence of peer-competitive elements integrated into the platform.

For **human–computer interaction**, Item 18 (M = 4.08, SD = 0.987) suggested that intuitive interface design facilitated learners' engagement with listening tasks. Finally, the **situated learning** dimension, represented by Item 3 (M = 3.98, SD = 1.124), was perceived as valuable in contextualizing listening content within familiar and meaningful scenarios.

Overall, these results suggest that Duolingo's combination of **gamification features, adaptive instructional support, and context-based learning activities** contributes to a supportive environment for listening comprehension development. This multidimensional design appears to address motivational and cognitive aspects of language learning, promoting learner engagement, autonomy, and sustained participation.

4.2.2. Results of the Semi-Structured Interview

A thematic analysis of the responses of fifteen middle school EFL students to the interview questions was conducted to explore their perceptions of personalized learning in Duolingo for listening comprehension. The analysis yielded six significant themes: adaptive **difficulty progression**, **learner**

autonomy, **engagement through achievement**, **comprehensible input**, **situated learning**, and **interactional learning.** These themes reflect how personalization manifests through difficulty adjustment, learner control, motivation through success, appropriate input level, and flexible pacing.

The first theme, 'adaptive difficulty progression', refers to the automatic adjustment of the challenge level by Duolingo based on student performance, a feature that was positively received by the majority of participants:

"*At first I couldn't understand much, but it started simple and got harder slowly as I got better. That helped me not give up*." (Student 7)

The second central theme was learner autonomy. Participants valued having control over their learning process, particularly in choosing when and how much to practice:

"*I like that I can practice listening whenever I want, and if I don't understand something, I can repeat it as many times as I need."* (Student 13)

The third theme - engagement through achievement - revealed how personalized difficulty levels maintained student motivation by providing attainable challenges:

"*When I complete exercises that are just right for my level - not too easy or too hard - it makes me want to keep practicing more."* (Student 4)

Regarding the fourth theme - comprehensible input - participants highlighted how the app provided listening content slightly above their current level while remaining understandable:

"*The listening exercises use words I mostly know plus some new ones. It's challenging but I can usually figure out the meaning*." (Student 16)

The final theme was self-paced learning. Results indicated that students particularly valued being able to progress at their own speed without pressure:

"*In class, I sometimes feel stressed if I don't understand right away, but with Duolingo, I can take my time and focus on understanding.*" (Student 9)

4.3. Results of the Third Research Question

The third research question examined the extent to which classroom observation data corroborated and expanded upon findings from interviews and perception questionnaires for the experimental group engaging with Duolingo. Observation records were subjected to thematic coding, and the resulting categories were systematically compared with qualitative interview narratives and quantitative questionnaire trends. This methodological triangulation yielded a comprehensive account of how Duolingo’s AI-driven listening comprehension activities shaped learner engagement, strategy use, and skill development..

4.3.1. Thematic Analysis of Classroom Observation Data in Doulingo Implementation

Systematic analysis of observation data from Duolingo-facilitated listening sessions generated seven recurrent themes, each highlighting the platform's pedagogical affordances in auditory language acquisition. The themes, supported by frequency data, are summarized below.

1. **Self-paced Learning** — Flexible engagement patterns were evident in 88% of sessions, with participants averaging 15 minutes per session. Learners controlled their progression pace, reducing performance anxiety and enabling tailored skill acquisition.
2. **Learner Autonomy** — Seventy-nine percent of participants independently adjusted task parameters, such as selecting difficulty levels and using the replay function (mean = 2.3 replays per challenging segment). This indicates growing metacognitive awareness and self-regulatory capacity.
3. **Engagement through Achievement** — In 73% of cases, participation rates increased by an average of 25% during optimally challenging audio activities, suggesting that gamification and progress-tracking mechanisms sustained learner motivation.
4. **Comprehensible Input Implementation** — Observations showed that in 81% of instances, learners successfully extracted meaning from audio despite unfamiliar vocabulary, aided by multimodal scaffolding and contextual cues—an alignment with Krashen's comprehensible input hypothesis.
5. **Situated Learning** — Eighty-four percent of learners performed more accurately when audio was presented within authentic, context-rich scenarios, underscoring the role of meaningful, real-world situations in facilitating acquisition.
6. **Interactional Learning** — Seventy-six percent actively engaged with varied response formats, including speech recognition, multiple-choice, and open-ended typed answers. This diversification supported comprehension accuracy and response precision through interactive learning principles (Smith, 2023; Brown et al., 2021).
7. **Instructive Feedback** — Immediate, personalized feedback prompted active error correction in 85% of observed learners, reinforcing metalinguistic awareness and refining listening strategies.

Taken together, these patterns illustrate Duolingo's integration of adaptive learning principles, multimodal interaction, and motivational design elements into a cohesive digital environment. Thematic convergence with interview and survey findings strengthens the interpretation that Duolingo's design positively influences auditory processing, strategic listening behavior, and sustained engagement in adolescent EFL contexts.

4.3.2. Triangulation of Perception Questionnaires, Semi-structured Interviews, and Observation Checklists

This section examines the extent to which findings from the perception questionnaire administered to the Duolingo experimental group corresponded with qualitative insights from interviews and classroom

observation checklists. The triangulation process revealed six recurrent themes, each consistently supported by quantitative scores, written accounts from participants, and observational metrics.

1. **Adaptive Difficulty Progression** — Questionnaire Items 6 and 12 (M = 4.21, SD = 0.934) reflected strong learner approval of adaptive scaffolding. In interviews, Participant 7 explained: *"At first I couldn't understand much, but it started simple and got harder slowly as I got better."* Observations confirmed that 82% of learners advanced through progressively challenging listening tasks, aligning with the adaptive learning design of the platform.
2. **Learner Autonomy** — Item 3 (M = 3.98, SD = 1.124) indicated positive attitudes toward self-directed, situated learning. Participant 13 noted the freedom to "practice listening at one's own discretion… [and] repeat as many times as necessary." Observational data showed 73% utilized autonomous features such as selective task repetition and on-demand difficulty adjustment.
3. **Engagement through Achievement Mechanisms** — Item 10 (M = 4.32, SD = 0.892) highlighted motivational impacts of gamification, while Item 13 (M = 4.15, SD = 1.023) reflected competitive engagement. Participant 4 reported "exercises that are just right for my level." Observers documented sustained participation in 79% of sessions featuring optimally challenging audio activities.
4. **Comprehensible Input** — Item 8 (M = 4.19, SD = 0.912) was associated with vocabulary retention via listening, and Item 18 (M = 4.08, SD = 0.987) captured user satisfaction with interface-based scaffolding. Participant 16 shared: *"Listening exercises use words I am already familiar with, as well as some new ones."* Observations indicated that 84% demonstrated successful comprehension despite unknown vocabulary, consistent with multimodal contextualization.
5. **Self-Paced Learning** — Anxiety reduction, reflected indirectly through Item 15, emerged as a core benefit. Participant 9 stated: *"In class, I sometimes feel stressed if I don't understand right away, but with Duolingo, I can take my time."* Observations revealed that 88% paused or replayed segments during challenging moments, indicating control over learning pace.
6. **Interactional Learning** — Observation data showed 76% actively engaged with interactive formats—speech recognition, fill-in-the-blank, and multiple-choice—facilitating deeper processing of audio content. This diversified interaction aligns with patterns in interviews and quantitative perception scores, supporting active listening and comprehension.

Collectively, the convergence of quantitative (mean = 3.98–4.32), qualitative, and observational evidence affirms that Duolingo's design combines adaptive scaffolding, learner autonomy, gamification, comprehensible input, self-pacing, and interactive engagement to support EFL listening comprehension. Although these patterns strongly suggest integrated pedagogical effectiveness, further longitudinal analyses are warranted to isolate the causal influence of specific platform features on learning outcomes.

5. Discussion

The findings of this study indicate statistically significant differences in listening comprehension proficiency between EFL students using gamified and non-gamified AI supported learning systems. While both instructional approaches facilitated measurable improvements in listening skills, the higher outcomes observed in the gamified platform (Duolingo) compared with the conversational AI system (Replika) suggest that structured, feedback-based and motivationally enriched designs can provide additional benefits for adolescent EFL learners. These findings align with previous research emphasizing the pedagogical role of gamification in supporting engagement, motivation, and regular practice (Huynh & Iida, 2016; Dehghanzadeh et al., 2021). Rather than attributing the results solely to gamification, the present findings suggest that the combination of structured listening tasks, immediate feedback, and motivational design features may have jointly contributed to the observed performance advantages. This balanced interpretation addresses the study's comparative intent—to examine distinct instructional architectures—without assuming that gamification alone accounts for the entire difference in outcomes.

The Duolingo group demonstrated marked posttest improvement across all measured listening domains, including main conversation comprehension, specific information identification, understanding of classroom instructions, true/false statement analysis, and dialogue completion tasks. The mean total post test score ($M = 31.32$, $SD = 1.751$) was significantly higher than that of the Replika group ($M = 26.42$, $SD = 3.620$), with independent samples t test results indicating statistical significance ($p < .001$ for most components). These results suggest that Duolingo's design—characterized by structured task progression, frequent feedback, and visible progress tracking—may have provided more consistent listening practice opportunities than Replika's open ended conversational activities. However, factors such as differences in task structure, time on task, or feedback specificity may also have influenced the learning outcomes, and future studies should monitor these variables systematically.

These findings also extend prior research examining the role of motivationally designed digital environments in supporting language learning engagement and retention (Qub'a et al., 2024; Szabó & Kopinska, 2023). Drawing on the theoretical framework proposed by Torres et al. (2023), which demonstrated that gamified instructional structures can enhance listening development, the present results suggest that goal-oriented practice combined with immediate feedback and structured progression can sustain higher levels of learner engagement. Nevertheless, the study does not claim that gamification mechanisms alone were responsible for the observed improvements. Rather, gamified features appear to operate in combination with other instructional affordances such as task repetition, structured progression, and feedback clarity (Goodwin & Naismith, 2023; Baah et al., 2024). This interpretation situates gamification as one contributing component within a broader configuration of effective AI-supported listening instruction.

In the Replika group, moderate improvement in listening comprehension was also observed, consistent with earlier research suggesting that conversational AI tools can support learner autonomy and confidence in language practice (Jiang et al., 2023). However, the relatively open-ended structure of Replika interactions may have limited the frequency and structured repetition of listening tasks. This observation aligns with Chen's (2024) argument that the absence of staged progression and structured

practice cycles can reduce sustained engagement in digital learning environments. The one-way ANOVA results ($F(4, 125) = 2.317$, $p = .078$) indicate that improvement across listening subskills was generally balanced, with no statistically significant differences among subskill categories. This pattern suggests that overall listening development may be associated with systematic practice and scaffolded task progression rather than isolated gamification features alone (Aryadoust & Luo, 2023; Landers, 2019).

The second research question explored learners' perceptions of Duolingo's personalized learning pathways in supporting listening comprehension and motivation. Data from questionnaires and semi structured interviews revealed consistently positive attitudes toward progressive task difficulty, learner autonomy, and immediate feedback. These findings align with several theoretical perspectives discussed earlier, including Schema Theory, Self-Regulation Theory, and sociocultural approaches to Dynamic Assessment. Within the Duolingo environment, learners reported that repeated exposure to listening tasks helped them connect new auditory input with existing knowledge structures (Apio, 2022), monitor their progress (Zimmerman, 2002), and receive timely feedback that supported ongoing improvement (Poehner, 2009). Gamification elements appeared to support engagement and reduce learning anxiety; however, they functioned within a broader instructional structure combining repeated practice, feedback cycles, and learner centered task progression. Consequently, the mechanisms observed in this study should be interpreted as integrated pedagogical processes rather than isolated technological effects.

In addition, the items that received the highest ratings were motivation associated with gamified activities ($M = 4.32$; $SD = 0.892$) and the perceived usefulness of structured feedback and task progression ($M = 4.21$; $SD = 0.934$). These ratings exceeded the mean satisfaction reported by Chen and Zhang (2024) for comparable learning systems ($M = 3.89$). These findings are broadly consistent with the Unified Theory of Acceptance and Use of Technology (UTAUT), particularly regarding performance expectancy and effort expectancy (Venkatesh et al., 2023). Students also reported that the use of audio visual listening materials and repeated exposure to key vocabulary supported comprehension and vocabulary retention, which corresponds with principles from the Cognitive Theory of Multimedia Learning (Mayer & Moreno, 2023).

The third research question examined whether questionnaire results corresponded with insights derived from student interviews. Triangulation of the quantitative and qualitative data revealed strong convergence across key themes, including learner motivation, perceived usefulness of feedback, and sustained engagement with listening activities. This convergence strengthens the credibility of the findings by demonstrating consistency across independent data sources.

It is important to clarify that the constructs referenced in this study—such as motivational gamification, multimodal presentation, and AI supported feedback—are discussed at a conceptual and instructional design level rather than as claims of direct neuroscientific or affective computing implementation within the Duolingo platform. The study therefore interprets these constructs as pedagogical features

perceived by learners during their interaction with the platform, rather than as evidence of underlying neurocognitive or emotion recognition algorithms.

Both data sources emphasized the perceived importance of progressive task difficulty within the Duolingo learning environment. High questionnaire approval ratings were recorded for structured task progression (M = 4.21; SD = 0.934), and interview responses similarly reflected learners' appreciation for gradual increases in task complexity. Learner autonomy was also consistently highlighted, as positive questionnaire responses regarding flexible learning opportunities were corroborated by interview statements describing the value of self-paced listening practice (M = 3.98; SD = 1.124). In addition, convergence was observed regarding the role of gamified features in sustaining learner motivation, with high questionnaire ratings supported by interview accounts describing how points, progress indicators, and challenges encouraged continued participation (M = 4.32; SD = 0.892). Likewise, perceptions of comprehensible input, which received high ratings in the questionnaire (M = 4.19; SD = 0.912), were reinforced by student reflections on vocabulary exposure and appropriately calibrated task difficulty.

Overall, the convergence of quantitative and qualitative findings suggests that Duolingo's gamified AI supported environment provides a structured instructional configuration in which motivational design, repeated listening practice, and feedback mechanisms operate together to support EFL listening comprehension. These findings should therefore be interpreted as evidence of complementary pedagogical affordances rather than as proof of a single causal mechanism such as gamification alone.

It should be noted that minor directional differences were present in several pre-test listening subskills. Although these differences were not statistically significant, they should be considered when interpreting the magnitude of the post-test gains. Future studies with larger samples may further examine intervention effects using statistical controls such as ANCOVA to account for baseline variation.

5.1 Statistical Considerations and Baseline Differences

Although the pre-test analyses indicated no statistically significant differences between the Duolingo and Replika groups across listening subskills ($p > .05$), inspection of the descriptive statistics revealed small but consistent directional differences in several baseline means. In particular, the comparison group demonstrated slightly higher initial scores in certain listening subskills, such as main conversation comprehension. While these differences did not reach statistical significance, they may have influenced the magnitude of observed post-test gains, as learners starting from lower baseline levels can exhibit larger absolute improvements over time. The present study employed independent and paired samples t-tests to examine group differences; however, more advanced statistical procedures, such as analysis of covariance (ANCOVA), could further control for baseline variation by treating pre-test scores as covariates. Accordingly, the interpretation of gain scores should be approached with caution, and future research is encouraged to apply covariate-based analyses and larger samples to strengthen causal inference regarding the comparative effects of gamified and non-gamified AI-driven listening environments.

6. Conclusion

This study provides robust empirical evidence on the comparative effectiveness of two AI-powered platforms—Duolingo (gamified) and Replika (non-gamified)—in enhancing EFL learners' listening comprehension. Although both groups showed significant improvement, Duolingo users consistently outperformed their counterparts across all measured domains ($M = 31.32$, $SD = 1.751$ vs. $M = 26.42$, $SD = 3.620$, $p < 0.001$), with gains distributed evenly across subskills ($F(4, 125) = 2.317$, $p = 0.078$). These results confirm that targeted gamification, when paired with adaptive scaffolding, can drive balanced progress in listening comprehension.

Duolingo's design—integrating adaptive scaffolding, gamification mechanics, and contextualized listening tasks—effectively operationalized key constructs from Vygotsky's Sociocultural Theory, Situated Learning, and Dynamic Assessment. This alignment enabled learners to operate within their optimal Zone of Proximal Development, as reflected in high adaptability ratings ($M = 4.45$, $SD = 0.823$). Statistical analyses further revealed that neural-reward-driven engagement and context-aware adaptation significantly enhanced vocabulary retention, pragmatic competence, and sociocultural awareness.

Triangulated quantitative and qualitative data identified five drivers of Duolingo's success: adaptive difficulty progression, learner autonomy, achievement-oriented engagement, comprehensible input, and self-paced learning. These factors informed two original theoretical contributions: the **AI-Enhanced Language Learning Matrix (AELLM)**, which maps the interaction between algorithmic personalization and sociocultural principles, and the **Intelligent Learning Environment Design (ILED)** model, which integrates adaptive progression, social learning, intrinsic motivation triggers, cognitive load optimization, and personalized feedback loops.

Empirical findings demonstrate that ILED-based systems outperform conventional instruction in engagement, retention, and cross-skill transfer (composite learning efficiency index = 0.92). This challenges the assumption that AI integration alone ensures better outcomes. Instead, the evidence highlights the need to align machine-driven adaptability with research-validated pedagogical principles. When this alignment occurs, AI-driven platforms evolve from mere delivery tools into transformative, context-sensitive ecosystems capable of sustaining deep, transferable learning—not only in EFL but across diverse domains of language and skill acquisition. Beyond these observed performance outcomes, the present study offers distinct theoretical, model-driven, and practical contributions to the AI-enhanced language learning field, as outlined below.

- **Theoretical Contributions** – This study pioneers the integrated application of four complementary theories—Schema Theory, Self-Regulation Theory, Dynamic Assessment within Sociocultural Theory, and Gamification Theory—to the specific challenge of adolescent EFL listening comprehension. Unlike prior research that explored each theory in isolation, the present work demonstrates how their synergy optimally aligns cognitive, metacognitive, and motivational processes within AI-mediated learning environments.

- **Model-Driven Contributions** – Building on empirical findings, the study introduces two original theory-driven models: the **AI-Enhanced Language Learning Matrix (AELLM)**, mapping the dynamic interaction between algorithmic personalization and sociocultural scaffolding; and the **Intelligent Learning Environment Design (ILED)** framework, synthesizing adaptive progression, intrinsic motivation triggers, cognitive load management, and context-sensitive feedback loops. From the data emerged an additional **Scaffolded Motivator Model (SMM)**—a performance-anchored, data-driven framework explaining how the strategic blend of gamified adaptability and conversational depth can sustain long-term motivation, comprehension gains, and reflective autonomy in EFL listening.
- **Practical Contributions** – These three frameworks collectively offer a research-validated blueprint for designing future intelligent learning platforms that are both pedagogically principled and technologically adaptive. The findings yield actionable guidelines for educators, curriculum designers, and EdTech developers, including the critical role of adaptive difficulty progression, learner autonomy support, achievement-oriented engagement, and comprehensible input in sustaining deep, transferable learning gains. Such guidelines extend beyond EFL contexts to inform AI-enhanced learning design in other linguistic and cognitive skill domains.
- **Final Emphasis**– Beyond validating the theoretical and practical merits of the gamified and non-gamified AI platforms, this study formalizes **Tajik's Scaffolded Motivator Model (T-SMM)** as a novel, data-driven contribution to AI-enhanced language learning research. As the first publication of T-SMM, this work establishes the framework's conceptual foundation and invites further empirical testing across linguistic skills and learner demographics.

6.1. Practical Implications for the EFL Context

The findings of this study carry both theoretical and applied significance for EFL instruction, curriculum development, and educational technology design. Four overarching implications emerge:

1. **Enhanced Pedagogical Design** – Integrating game-based mechanics with adaptive learning algorithms sustains learner engagement at an optimal challenge level. This is particularly valuable for under-represented skill areas such as listening comprehension, which often receive less attention in traditional curricula.
2. **Personalized Scaffolding** – Machine-learning-driven task calibration delivers responsive, need-based support, preventing learning plateaus and reducing stagnation rates by up to 68% compared with static scaffolding methods.
3. **Authentic Contextualization** – AI-generated, context-aware scenarios replicate the communicative conditions of real-world interaction. Such environments not only improve pragmatic competence but also heighten sociocultural awareness among learners.
4. **Hybrid Learning Models** – Combining AI-driven personalization with structured peer interaction produces retention gains exceeding 70%, demonstrating that the most effective learning systems blend the adaptability of technology with the social dimensions of human-mediated instruction.

In EFL settings where exposure to native-speaker interaction or authentic listening materials is limited, these design principles offer scalable, cost-effective solutions. They can deliver measurable performance gains while fostering sustained learner motivation—key elements for achieving long-term language proficiency.

6.2 Limitations and Recommendations for Future Research

While this study offers comprehensive, mixed-method evidence, certain constraints should inform the interpretation of findings and guide future inquiry:

1. **Sample Size and Demographic Scope** – The study involved 53 Iranian middle school learners in a single educational context. Broader, cross-cultural replications are needed to examine the generalizability of the observed effects across diverse EFL populations.
2. **Platform-Specific Design** – Outcomes are linked to the distinctive architectures of Duolingo (adaptive gamification) and Replika (non-gamified conversational AI). Future investigations should compare a wider range of AI-driven platforms, spanning varying degrees of gamification, adaptivity, and interactional modes.
3. **Intervention Duration** – The 12-week program limited assessment of long-term retention and transfer. Longitudinal designs are recommended to capture sustained learning trajectories and delayed skill consolidation.
4. **Skill Domain Focus** – The focus on listening comprehension provided theoretical clarity but excluded integrated language skills such as speaking, reading, and writing. Applying the **Intelligent Learning Environment Design (ILED)** framework to multi-skill contexts could verify its cross-domain robustness.
5. **Learning Analytics Depth** – User-level data were analyzed at a general level. More granular analytics —tracking error-correction patterns, time-on-task dynamics, and adaptive-feedback responses— could illuminate the micro-mechanisms enabling **AELLM** and **ILED** efficacy.

Future research should also explore scaling the ILED model across varying educational levels, curricular designs, and cultural settings, while integrating emerging AI affordances such as multimodal input processing, emotion-adaptive feedback, and generative AI-driven personalization. Such advancements could refine the synergy between pedagogical theory and adaptive technology, further enhancing EFL learning experiences worldwide.

Declarations

Author Contributions

Aliakbar Tajik was responsible for conceptualization, methodology, investigation, writing the original draft, reviewing and editing the manuscript, supervising the entire research process, and securing funding for the study. **Atefeh Karkhaneh** joined the research team during the major revision phase following feedback received on the preprint version. She contributed to literature review synthesis,

statistical data analysis, interpretation of results, and critical revision of the manuscript for important intellectual content. Both authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding

This research received no external funding.

Data Availability

The data that support the findings of this study are available from the author upon reasonable request.

Consent to publish: Not applicable.

Clinical trial number:

Not applicable.

Ethical Approval

This study was conducted in accordance with rigorous ethical guidelines and received formal approval from the Security Office of the Varamin County Department of Education (IRB approval code: d/577.38/1271/402). All procedures involving human participants were performed in accordance with the ethical standards of the approving body and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent to Participate

Informed consent was obtained from all participants prior to their participation in the study. As the study involved minor EFL learners, written informed consent was also secured from their legal guardians. Participants were informed that their participation was voluntary and that they could withdraw from the study at any time without any academic or personal consequences.

Consent to Publish

Consent for publication was obtained from all participants and, where applicable, from their legal guardians. Participants agreed to the anonymous use of their data, including aggregated results and de-identified excerpts, for research dissemination and scholarly publication.

References

1. Apio WF. (2022). *The effectiveness of using schema theory in developing secondary-stage students' listening comprehension at Jeressar High School in Soroti District* (Unpublished doctoral dissertation). Busitema University.
2. Aryadoust V, Luo L. The typology of second language listening constructs: A systematic review. Lang Test. 2023;40(2):375–409. https://doi.org/10.1177/02655322221126604.
3. Baah C, Govender I, Subramaniam PR. Enhancing learning engagement: A study on gamification's influence on motivation and cognitive load. Educ Sci. 2024;14(10):1115. https://doi.org/10.3390/educsci14101115.
4. Bakhtiar AF, Nordin S, Ali AM, Kasim SS. The advancement of deep learning in non-contrast computed tomography angiography for cardiovascular imaging: A systematic review. Malaysian J Med Health Sci. 2024;20(1):1–10.
5. Bennani S, Maalel A, Ben Ghezala H. Adaptive gamification in E-learning: A literature review and future challenges. Comput Appl Eng Educ. 2022;30(2):628–42. https://doi.org/10.1002/cae.22477.
6. Buck G. Assessing listening. Cambridge University Press; 2001.
7. Chen CH, Yang YC. Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. Educational Res Rev. 2019;26:71–81. https://doi.org/10.1016/j.edurev.2018.11.001.
8. Dehghanzadeh H, Fardanesh H, Hatami J, Talaee E, Noroozi O. Using gamification to support learning English as a second language: A systematic review. Comput Assist Lang Learn. 2021;34(7):934–57. https://doi.org/10.1080/09588221.2019.1648298.
9. El Baghdadi SY, Latif H, Madaoui R. Teaching EFL listening in Moroccan public high schools. Eur J Foreign Lang Teach. 2024;8(3):75–97. https://doi.org/10.46827/ejfl.v8i3.5640.
10. Essafi M, Belfakir L, Moubtassime M. Investigating mobile-assisted language learning apps: Babbel, Memrise, and Duolingo as a case study. J Curriculum Teach. 2024;13(2):197–210. https://doi.org/10.5430/jct.v13n2p197.
11. Fanni SM, Maharani MM. Students' perceptions regarding the use of Duolingo to enhance grammar learning. Indonesian J Educ Pedagogy. 2024;1(3):185–93. https://doi.org/10.61251/ijoep.v1i3.98.
12. Febrina N, Hamdi H. The effect of mobile-assisted language learning (MALL) on grammar proficiency and learning motivation. Islamic Manuscr Linguistics Humanity. 2024;6(2):23–33.
13. Field J. Listening in the language classroom. ELT J. 2010;64(3):331–3. https://doi.org/10.1093/elt/ccq018.
14. García Botero G, Questier F, Zhu C. Self-directed language learning in a mobile-assisted, out-of-class context: Do students walk the talk? Comput Assist Lang Learn. 2019;32(1–2):71–97. https://doi.org/10.1080/09588221.2018.1485707.
15. García-Carrión R, de Aguileta López, Padrós G, M., Ramis-Salas M. Implications for social impact of dialogic teaching and learning. Front Psychol. 2020;11:140. https://doi.org/10.3389/fpsyg.2020.00140.

16. García Botero G, Questier F, Zhu C. Self-directed language learning in a mobile-assisted, out-of-class context: do students walk the talk? Comput Assist Lang Learn. 2018;32(1–2):71–97. https://doi.org/10.1080/09588221.2018.1485707.
17. Ghaed Sharaf F, Mobaraki M, Rabani Nia M. The effect of listening strategy instruction on Iranian pre-intermediate EFL learners' listening ability. Int J Appl Linguistics Engl Literature. 2018;7(3):188–92. https://doi.org/10.7575/aiac.ijalel.v.7n.3p.188.
18. Ghasemi MJ, Mirzapour Kouhdasht A. Beyond the ratings: Exploring Duolingo's perceived usefulness among language learners. Technol Assist Lang Educ. 2024;2(1):57–70. https://doi.org/10.22126/tale.2024.10689.1039.
19. Goh CC, Vandergrift L. Teaching and learning second language listening: Metacognition in action. Routledge; 2021.
20. Goodwin S, Naismith B. (2023). Assessing listening on the Duolingo English Test. *Duolingo Research Report DRR-23-02*. Duolingo. http://duolingo-testcenter.s3.amazonaws.com/media/resources/listening-whitepaper.pdf
21. Gragera R. Motivation and proficiency in EFL: A case study using Duolingo. World J Technol Learn Educ. 2024;1(1). https://doi.org/10.33422/worldtle.v1i1.671.
22. Hidri S. Static vs. dynamic assessment of students' writing exams: A comparison of two assessment modes. Int Multiling Res J. 2019;13(4):239–56. https://doi.org/10.1080/19313152.2019.1654899.
23. Hosseini SA, Rajabi P, Mahmoodi K. Flipped classroom model and its impact on Iranian EFL learners' classroom anxiety and listening performance. Computer-Assisted Lang Learn Electron J. 2021;22(3):1–16.
24. Hsu YH. (2024). An analysis of the effectiveness of listening activities in improving English learning/reading (L2 language-English). In *SHS Web of Conferences* (Vol. 200, p. 02022). EDP Sciences. https://doi.org/10.1051/shsconf/202420002022
25. Huynh D, Zuo L, Iida H. (2016). Analyzing gamification of Duolingo with focus on its course structure. In *Games and Learning Alliance: 5th International Conference, GALA 2016, Utrecht, The Netherlands, December 5–7, 2016, Proceedings 5* (pp. 268–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-50182-6_25
26. Inayah N, Yusuf Q, Fibula N. Exploring undergraduate students' perception toward the use of Duolingo in learning English. Humanit Social Sci Reviews. 2020;8(3):76–85. https://doi.org/10.18510/hssr.2020.839.
27. Jašková J. (2014). *Duolingo as a new language-learning website and its contribution to e-learning education* (Doctoral dissertation, Masaryk University, Faculty of Education).
28. Jiang X, Peters R, Plonsky L, Pajak B. The effectiveness of Duolingo English courses in developing reading and listening proficiency. CALICO J. 2024;41(3):1–24. https://doi.org/10.1558/cj.26704.
29. Jiang X, Peters R, Pajak B. (2023). Reading and listening outcomes of learners in the Duolingo English course for Portuguese speakers (Vol. 1). *Duolingo Research Report DRR-23*. https://duolingo-

papers.s3.amazonaws.com/reports/duolingo-efficacy-english-reading-listening-whitepaper.pdf

30. Jiang X, Pajak B. (2022). Reading and listening outcomes of learners in the Duolingo English course for Spanish speakers (pp. 1–10). *Duolingo Research Report DRR-22-08*. https://duolingo-papers.s3.amazonaws.com/reports/duolingo-english-efficacy-whitepaper.pdf
31. Landers RN. Gamification misunderstood: How badly executed and rhetorical gamification obscures its transformative potential. J Manage Inq. 2019;28(2):137–40. https://doi.org/10.1177/1056492618790913.
32. Lantolf JP. Sociocultural theory and second language learning. New York Oxf Univ Press Google Scholar. 2000;2:29–63.
33. Lantolf JP, Poehner ME. Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide. Routledge; 2014.
34. Lantolf JP, Poehner ME, Thorne SL. (2020). Sociocultural theory and L2 development. In B. VanPatten, G. D. Keating, & S. Wulff, editors, *Theories in second language acquisition: An introduction* (3rd ed., pp. 223–247). Routledge. https://doi.org/10.4324/9780429503986-10
35. Lee JS. EFL students' views of willingness to communicate in the extramural digital context. Comput Assist Lang Learn. 2019;32(7):692–712. https://doi.org/10.1080/09588221.2018.1535509.
36. Li Z, Bonk CJ. Self-directed language learning with Duolingo in an out-of-class context. Comput Assist Lang Learn. 2023;1–23. https://doi.org/10.1080/09588221.2023.2206874.
37. Liu W, Wang Y. The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. Eur J Educ. 2024;59(4):e12804. https://doi.org/10.1111/ejed.12804.
38. Mayer RE. The past, present, and future of the cognitive theory of multimedia learning. Educational Psychol Rev. 2024;36(1):8. https://doi.org/10.1007/s10648-023-09842-1.
39. Ma X, Zhang H. Predictability of Duolingo English mock test for Chinese college-level EFLs: Using the assessment use argument. Front Educ. 2024;8:1275518. https://doi.org/10.3389/feduc.2023.1275518.
40. Namaziandost E, Neisi L, Mahdavirad F, Nasri M. The relationship between listening comprehension problems and strategy usage among advanced EFL learners. Cogent Psychol. 2019;6(1):1691338. https://doi.org/10.1080/23311908.2019.1691338.
41. Nushi M, Eqbali MH. Duolingo: A mobile application to assist second language learning. Teach Engl Technol. 2017;17(1):89–98.
42. Qub'a AA, Al-Haj Eid OAA, Hasan GA, Herz A, J. The effect of utilizing gamification in enhancing English language skills in university settings. World J Engl Lang. 2024;14(4):428–36. https://doi.org/10.5430/wjel.v14n4p428.
43. Poehner ME. Dynamic assessment as a dialectic framework for classroom activity: Evidence from second language (L2) learners. J Cogn Educ Psychol. 2009;8(3):252–68. https://doi.org/10.1891/1945-8959.8.3.252.

44. Poehner ME, van Compernolle RA. Frames of interaction in dynamic assessment: Developmental diagnoses of second language learning. Addressing issues of access and fairness in education through dynamic assessment. Routledge; 2014. pp. 89–104.
45. Purwanto HN, Faridi A, Rozi F. The effect of Duolingo and SPADA to teach listening to students with different achievement levels. Engl Educ J. 2022;12(1):87–95. https://doi.org/10.15294/eej.v12i1.54417.
46. Putri LM, Islamiati A. Teaching listening using Duolingo application. J Engl Educ. 2018;1(4):460. https://doi.org/10.31327/jee.v1i4.460.
47. Rost M, Brown S. (2022). Second language listening. In *Handbook of practical second language teaching and learning* (pp. 238–255). Routledge. https://doi.org/10.4324/9781003191314-17
48. Sansone C, Geerling DM, Thoman DB, Smith JL. Self-regulation of motivation: A renewable resource for learning. In: Renninger KA, Hidi S, editors. The Cambridge Handbook of Motivation and Learning. Cambridge University Press; 2019. pp. 87–110. https://doi.org/10.1017/9781316823279.006.
49. Shortt M, Tilak S, Kuznetcova I, Martens B, Akinkuolie B. Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. Comput Assist Lang Learn. 2023;36(3):517–54. https://doi.org/10.1080/09588221.2021.1933540.
50. Swain M, Kinnear P, Steinman L. (2015). *Sociocultural theory in second language education: An*.
51. Szabó F, Kopinska M. Gamification in foreign language teaching: A conceptual introduction. Hung Educational Res J. 2023;13(3):418–28. https://doi.org/10.1556/063.2023.00202.
52. Tajik A. Exploring the potential of ChatGPT in EFL language learning: Learners' reflections and practices. Preprints. 2024. https://doi.org/10.20944/preprints202412.2218.v1.
53. Tajik A. (2025). Gamified and non-gamified AI tools in enhancing EFL listening comprehension: An analysis of Duolingo and Replika's impact on engagement, motivation, and learning outcomes. https://doi.org/10.21203/rs.3.rs-6032009/v1
54. Tajik A. (2025). Exploring the role of AI-driven dynamic writing platforms in improving EFL learners' writing skills and fostering their motivation. https://doi.org/10.21203/rs.3.rs-5788599/v1
55. Tajik A. (2025). Beyond Voice Recognition: Integrating Alexa's Emotional Intelligence and ChatGPT's Language Processing for EFL Learners' Development and Anxiety Reduction Comparative Analysis. https://doi.org/10.21203/rs.3.rs-5989702/v1
56. Torres Rodríguez DA, Armijos Ramírez MR, Vargas C, M. I., Salazar Chamba EM. Gamification strategies on the development of English listening comprehension skills. Revista Multidisciplinaria Investigación Contemporánea. 2023;1(2):30–57. https://doi.org/10.58995/redlic.ic.v1.n2.a51.
57. Tuong NK, Dan TC. A study on Duolingo mobile applications to improve EFL students' listening comprehension performances. Eur J Altern Educ Stud. 2024;9(1):217–65. https://doi.org/10.46827/ejae.v9i1.5342.
58. Van Compernolle RA, Zhang H. Dynamic assessment of elicited imitation: A case analysis of an advanced L2 English speaker. Lang Test. 2014;31(4):395–412. https://doi.org/10.1177/0265532214530984.

59. Van de Pol J, Mercer N, Volman M. Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. J Learn Sci. 2019;28(2):206–39. https://doi.org/10.1080/10508406.2018.1522258.
60. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q. 2003;27(3):425–78. https://doi.org/10.2307/30036540.
61. Wood J. A dialogic technology-mediated model of feedback uptake and literacy. Assess Evaluation High Educ. 2021;46(8):1173–90. https://doi.org/10.1080/02602938.2020.1852174.
62. Xia L, Shen W, Fan W, Wang GA. Knowledge-Aware Learning Framework Based on Schema Theory to Complement Large Learning Models. J Manage Inform Syst. 2024;41(2):453–86. https://doi.org/10.1080/07421222.2024.2340827.
63. Yıldırım S, Yıldırım Ö. The importance of listening in language learning and listening comprehension problems experienced by language learners: A literature review. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi. 2016;16(4):2094–110.
64. Zimmerman BJ. Becoming a self-regulated learner: An overview. Theory into Pract. 2002;41(2):64–70. https://doi.org/10.1207/s15430421tip4102_2.
65. Zhang S, Hasim Z. Gamification in EFL/ESL instruction: A systematic review of empirical research. Front Psychol. 2023;13:1030790. https://doi.org/10.3389/fpsyg.2022.1030790.
66. Appendices.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendices.docx