Journal of Management Practices, Humanities and Social Sciences

Vol 10 Issue 6 pp. 906-924

https://doi.org/10.33152/jmphss-10.6.3



ORIGINAL CONTRIBUTION

EFL Speaking Performance in a Universal Design for Learning-Based Gamified Framework: A Quasi-Experimental Study

Suheyla Demirkol Orak[1], Mohammad Hamad Al-khresheh[2]*

[1]Foreign Languages Department, Faculty of Education, Fırat University, Elazığ, Türkiye

[2]Humanities and Social Research Center, Northern Border University, Arar 73213, Saudi Arabia

Abstract— This study examined whether a Universal Design for Learning (UDL)-based gamified framework was associated with differences in EFL speaking performance relative to conventional instruction. A quasi-experimental pretest–posttest design was employed with 42 undergraduate EFL learners enrolled in two intact classes at a public university in Turkey, with one intact class assigned to each instructional condition. One class received UDL-based gamified speaking instruction for 14 weeks, while the comparison class followed conventional speaking instruction under the same syllabus, instructional time, and assessment requirements. We assessed speaking performance using an analytic rubric covering fluency, grammatical accuracy, vocabulary range, and pronunciation, along with an overall speaking score. Two independent raters evaluated all performances, with inter-rater reliability coefficients ranging from .81 to .88. Post-test outcomes were analyzed using ANCOVA with corresponding pre-test scores entered as covariates, and Holm adjustment was applied to the four component-level comparisons. The experimental group achieved a significantly higher adjusted overall speaking score than the control group, $F(1, 39) = 26.88$, $p < .001$, partial $\eta^2 = .41$. At the component level, grammatical accuracy remained significant after correction, $p < .004$, partial $\eta^2 = .45$, as did pronunciation, $p = .042$, partial $\eta^2 = .15$. Fluency and vocabulary range showed higher adjusted means in the experimental group but did not remain significant after multiplicity correction. The findings indicate that integrating UDL principles with gamified instructional organization is associated with stronger overall speaking performance, while component-level responses differ rather than remaining uniform. The study extends UDL and gamification research into performance-based EFL speaking assessment and supports closer attention to how instructional design relates to distinct dimensions of oral performance.

Index Terms— Analytic speaking assessment, EFL speaking performance, Gamified instruction, Quasi-experimental design, Universal Design for Learning, University EFL context

Received: 5 June 2026; **Accepted:** 2 October 2026; **Published:** 6 November 2026



Introduction

Speaking is one of the most challenging skills in teaching English as a Foreign Language (EFL) because speaking requires learners to link their linguistic knowledge with language use in real-time settings. While speaking, learners must search for and retrieve appropriate vocabulary and maintain grammatical accuracy (Al-khresheh & Alruwaili, 2026; Derakhshan et al., 2016; Tavakoli et al., 2020; X. Yan et al., 2020). They must also make their speech intelligible and sustain communication while expressing themselves fluently in oral contexts. Given these requirements, teaching speaking in classroom settings is more complex, especially when learners differ in proficiency level, speaking anxiety, speaking ability,

*Corresponding Author: mohd_khresheh@yahoo.com

© 2026 The Author(s). All rights reserved.

processing speed, and willingness to use English in speaking tasks. Typically, conventional speaking exercises provide limited practice, such as tasks designed for all learners in the same way, within limited time frames, teacher-led activities, and very short speaking tasks. However, these traditional speaking practices do not meet learners' diverse needs in speaking classes (Abdullah et al., 2019; Namaziandost et al., 2019; Saed et al., 2021). The major challenge in teaching speaking, therefore, is not providing much speaking practice but designing and implementing effective speaking lessons in which all learners can participate and reach desired speaking competence while meeting their individual needs in speaking classes.

UDL provides a useful instructional basis for addressing this problem because it treats learner variability as an expected feature of instruction rather than an exception to a standard instructional pathway (Al-Azawei et al., 2016; Rusconi & Squillaci, 2023). Its relevance to EFL speaking lies in creating different routes into oral participation without changing the communicative goals learners are expected to achieve. Flexibility alone, though, does not resolve the need for sustained practice, progression, and continued engagement across speaking tasks. Gamification is particularly relevant here because game-informed structures can organize learners' movement through activities using goals, staged progression, feedback, challenges, and visible advancement (Manzano-León et al., 2021; S. Zhang & Hasim, 2023). UDL and gamification can thus address complementary aspects of the same pedagogical problem. UDL provides flexibility in how learners access and participate in speaking activities, whereas gamification structures progression, feedback, and continuity across those activities. In this framework, the two are not assumed to interact in ways that strengthen one another. Instead, each serves a different purpose within the same instructional design. UDL responds to differences in learners' readiness and participation, while gamification gives their speaking practice a clearer sense of progression and continuity. Integrating these functions into a single framework connects flexible participation with sustained communicative practice, which forms the main theoretical basis for their integration.

The empirical basis for this integration remains insufficiently developed. UDL research has largely focused on accessibility, inclusion, differentiation, and learner participation, leaving its relationship with directly assessed EFL speaking performance comparatively unclear (Bray et al., 2023; Ismailov & Chiu, 2022; Rusconi & Squillaci, 2023). Gamification research presents a different limitation. Although positive findings have been reported for motivation, engagement, participation, and language learning (Huang, 2021; Huseinović, 2023; Khaldi et al., 2023; Luo, 2023) increased willingness to participate does not necessarily translate into improved speaking performance. More importantly, speaking itself is multidimensional. An instructional framework associated with stronger overall performance may relate differently to fluency, grammatical accuracy, vocabulary range, and pronunciation because these dimensions place different demands on learners (Ginther et al., 2010; Ounis, 2017; Tavakoli et al., 2020). What remains unclear, then, is whether integrating UDL principles with gamified instructional organization is associated with measurable speaking performance relative to conventional instruction and whether any advantage extends consistently across the principal dimensions of oral performance. This question shifts the discussion from whether learners find UDL accessible or gamification engaging to the more demanding question of what learners can actually demonstrate in assessed oral performance.

The present study addresses this problem through a 14-week quasi-experimental comparison of UDL-based gamified speaking instruction and conventional instruction among university EFL learners. Rather than using engagement, motivation, or learner perceptions as indicators of instructional success, the study examines directly assessed speaking performance as an overall outcome and across fluency, grammatical accuracy, vocabulary range, and pronunciation. This distinction allows the study to test whether the integrated instructional framework is associated with stronger overall oral performance while also determining whether the pattern is consistent across its component dimensions. In doing so, the study extends UDL from access and participation to performance-oriented EFL instruction. It examines gamification as part of instructional organization rather than as a set of motivational additions to otherwise conventional teaching. The study thus provides a performance-based test of an integrated framework designed around two persistent requirements of EFL speaking pedagogy: accommodating learner variability and sustaining structured progression through communicative practice.

Literature Review

EFL speaking performance and instructional challenges

Speaking in a second or foreign language involves several interacting linguistic abilities that depend on one another during oral production. Therefore, speaking in a second language or foreign language cannot be reduced to a single ability. As second-language learners develop speaking fluency, temporal characteristics of speech, such as speed, continuity, pausing, and repair, help assess oral proficiency (Ginther et al., 2010). Different fluency measures, such as pause types, repair types, and local and global fluency, distinguish levels of speaking proficiency (Tavakoli et al., 2020). In addition to fluency, a speaker's lexical resources are an important factor in oral production because they determine how fully the speaker can convey meaning during communication (Janebi Enayat & Derakhshan, 2021). Pronunciation, another important factor in oral production, determines a speaker's intelligibility (Yenkimaleki & van Heuven, 2019). Grammatical accuracy, which several empirical studies have examined e.g., (Navidinia et al., 2019) is another important aspect of speaking because it shows the extent to which a speaker controls a language's linguistic structures during oral production. Therefore, fluency, grammatical accuracy, vocabulary range, and pronunciation need to be studied separately because development in one aspect does not automatically mean the same development in the others (Ounis, 2017; X. Yan et al.,

2020). From a research perspective, speaking test scores are often summed into an overall score that indicates a test taker's general speaking ability. However, this overall score provides no information about a test taker's strengths and weaknesses. Therefore, summing the scores of the different components of a speaking test provides a more detailed account of differences in the speaking ability of test takers from different backgrounds (Erdiana et al., 2019; Tuan, 2012).

Speaking is a complex skill that requires many opportunities to produce language, negotiate meaning, retrieve vocabulary, and maintain conversation during extended interaction. These opportunities are not always easily available for EFL learners who mostly interact with native speakers of the target language in the classroom. Thus, classroom interaction is the major source of speaking practice for EFL learners (Abdullah et al., 2019; Alrasheedi, 2020). Studies have associated interactive oral activities with the development of speaking ability (Derakhshan et al., 2016; Shen & Chiu, 2019). Studies have also shown that several linguistic, psychological, and instructional factors contribute to the difficulties learners encounter in developing L2 speaking skills (Shen & Chiu, 2019). To enable learners to engage in interactive oral activities, teachers need to organize classroom activities that give learners adequate opportunities to participate in spoken interactions. However, the extent to which learners participate in spoken interactions can differ from one learner to another, depending on factors such as proficiency, willingness to speak, and anxiety. Learners also differ in how confident and ready they are to communicate orally. Consequently, spoken interactions designed and delivered to meet all learners' needs are unlikely to be effective unless all learners have the opportunity to engage in relevant types of spoken interactions with adequate frequency and duration (Hanifa, 2018). A range of studies has identified anxiety and other negative emotions as key challenges that prevent learners from fully participating in classroom interaction (R. H. Chen, 2022). Furthermore, emotion regulation, through cognitive reappraisal or expressive suppression, plays a significant role in EFL contexts (Bourzeg, 2025). This means teachers need to create classroom conditions that let all learners participate freely in spoken interactions without fear of ridicule, and support them when they experience emotional distress while speaking.

The organization of practice is equally consequential. Speaking development requires repeated opportunities to use language for communication and to reconsider performance across successive attempts. Research on communicative and collaborative approaches indicates that extended interaction and active learner participation can support oral development (Afifah & Devana, 2020; Darmuki et al., 2018; Tsou, 2005). Feedback and learner attention to their own production can also support specific dimensions of speaking. For example, Navidinia et al. (2019) found that learners who recorded and transcribed their oral production and identified grammatical, phonological, and lexico-semantic errors subsequently produced fewer speaking errors than learners in the comparison condition. Technology-supported instruction also expands opportunities for oral practice and feedback. Studies report favorable speaking outcomes in technology-mediated settings where learners repeatedly engage with oral tasks and receive instructional support that differs from conventional classroom interaction (Bashori et al., 2021; Mulyadi et al., 2021; Tai, 2022). Al-khresheh (2025) provides more direct experimental evidence from AI-supported speaking practice, reporting effects on EFL learners' fluency and speaking anxiety. Qiao and Zhao (2023) likewise reported gains in speaking skills and speaking self-regulation following AI-based language instruction. These findings point to a broader instructional principle: speaking development depends partly on how the learning environment structures opportunities for oral production, interaction, feedback, and repeated performance.

However, even with sufficient speaking activities, learner variability remains a challenge. What works for one student may be too difficult for another. Fixed participation structures can also distribute speaking opportunities inequitably within a single class. Research has also begun to show correlations between AI-enhanced identity expression, growth mindset, language self-efficacy, and second-language grit, and their influence on learning in language education (M. A. Almayez et al., 2025; Yusuf, 2025). A range of comparative studies has shown differences between spoken language use in classes receiving conventional instruction and that in classes using task-based, collaborative, and technology-supported learning, as well as differences in the speaking that takes place within these different types of classrooms when there are different arrangements with respect to learning participation and practice (Asratie et al., 2023; Buckingham & Alpaslan, 2017; Nair & Yunus, 2021; Saed et al., 2021; Zhussupova & Shadiev, 2023). Learner autonomy and self-regulation are also influenced by the design of technology-supported language learning (Santhanasamy & Yunus, 2021). Thus, while it is indeed challenging to design lessons that are adequate for all students, the central pedagogical problem goes beyond adding more speaking activities to current EFL lessons, and this can be best addressed by designing an instructional framework that from the outset addresses and accommodates differences in learners' readiness to participate and engage in purposeful oral production with appropriate support and feedback, and that progresses gradually through a variety of different communicative tasks (Al-khresheh, 2025).

UDL and Gamified instruction in EFL

Varied student abilities and needs are inherent in all students learning in educational settings (Dobinson & Dockrell, 2021; Rusconi & Squillaci, 2023). UDL has increasingly been successfully applied in many EFL contexts through a proactive curricular approach (Quaglia, 2015). The three principles of UDL, namely multiple means of engagement, representation, and action or expression, are seen to support the provision of a flexible approach to student access and engagement in learning, as well as varying pathways for students to engage and express what they know and can do (Lintangsari & Emaliana, 2020). Moreover, several national and international studies have reported positive findings on academic, cognitive, and behavioral outcomes in educational settings where UDL was implemented (Al-Azawei et al., 2016). For example, applying UDL principles to

curriculum design for EFL learners can involve providing scaffolded materials, using multiple forms of representation, and giving learners different ways to express what they know (Rao & Meo, 2016). However, in practice, teachers do not always implement the three UDL design principles in a balanced way. For example, Bray et al. (2023) found that UDL design in the classrooms under investigation tended to focus more on representation than on engagement and action or expression. In the context of an asynchronous online EFL environment, for example, Ismailov and Chiu (2022) reported positive correlations between their implementation of UDL-design and two of the key tenets of the Self-Determination Theory, that is, learner autonomy as well as learners' perceived competence; however, they also reported that the third key tenet of the theory, that is, learners' sense of relatedness, was not well-supported in their online study.

In EFL contexts, UDL is increasingly viewed as a proactive approach to curriculum design. By viewing learner variability as a predictable aspect of human learning, rather than something to address after the fact through remediation, UDL can support design for difference in the classroom (Dobinson & Dockrell, 2021; Rusconi & Squillaci, 2023). Quaglia (2015) outlined three key principles for UDL: Multiple means of engagement, multiple means of representation, and multiple means of action or expression. In EFL teaching and learning, these principles matter because learners' linguistic capacity, communicative maturity, learning needs, and preferred modes of participation vary widely. Planning classroom-based learning with UDL in mind can support differentiated instruction and remove barriers to participation for students with a range of linguistic, cognitive, and sensory needs (Lintangsari & Emaliana, 2020). Implementing UDL across different educational contexts has been shown to improve accessibility and academic, cognitive, and behavioral outcomes (Al-Azawei et al., 2016). However, teachers do not always apply UDL's three principles equally. For example, a recent study found that teachers provided multiple means of representation more frequently than multiple means of engagement and multiple means of action or expression in their enacted UDL practices (Bray et al., 2023). In an asynchronous online EFL setting, UDL use was positively associated with learner autonomy and perceived competence. However, relatedness was supported less strongly than the other two key aspects of social presence (Ismailov & Chiu, 2022). These findings suggest that flexibility in access and participation represents an important part of instructional design, while sustained engagement and interaction require attention to how learning activities themselves are organized.

Gamified instruction addresses this latter issue by using game-informed elements in a structured way within learning activities. Evidence from EFL contexts links gamification to speaking-related outcomes, learner engagement, and participation, though results depend largely on instructional design and the functions assigned to individual game elements. S. Zhang and Hasim (2023) report generally favorable relationships between gamification and language-learning motivation, engagement, and learning outcomes. More specifically, Huseinović (2023) found improved speaking proficiency alongside greater learner participation in a gamified EFL setting. Gamified collaborative activities have also been associated with active participation and sustained involvement in group learning (Laura-De La Cruz et al., 2023), while Huang (2021) reports favorable language-learning outcomes when gamified activities are incorporated into classroom instruction. Related research extends this function beyond motivation. Z. Zhang and Crawford (2023) link gamified formative assessment to feedback and progression across learning activities, and Cabrera-Solano (2020) shows how digital game-based projects can provide structured opportunities for communicative language use. Gamification can thus serve an instructional function when goals, challenges, feedback, and progression are connected directly to the language activities learners are expected to perform.

The value of gamification therefore depends on how its elements relate to the activity's pedagogical goals. To add value, game elements should align with the activity's educational goals rather than being added to existing teaching practices (Ghai & Tandon, 2022). Khaldi et al. (2023) identified variability in gamification in terms of design and educational outcomes. While gamification has been found to increase learner motivation and participation, positive effects cannot be assumed simply from including elements such as points, rewards, competition, and progression (Luo, 2023). Not all gamification elements have equal value, so teachers need to select those most likely to work in their specific teaching context. Furthermore, while increased motivation and participation are desirable, they do not necessarily translate into improved oral proficiency. Gamified activities therefore have greater value when they facilitate purposeful language use and when feedback, task progression, and supporting structures are mapped to learners' language-learning goals.

Integrating UDL with gamification can offer a relevant, meaningful approach to teaching and learning. While both UDL and gamification aim to increase learner participation, they approach this goal differently (Al-Azawei et al., 2016; Rusconi & Squillaci, 2023). UDL aims to increase access, engagement, and expression by removing barriers to learning. It does this by providing multiple means of representation, expression, and engagement for learners with different levels of ability, learning style, and experience. By contrast, gamification structures learning to increase participation and engagement by setting goals, providing feedback, offering challenges, and enabling learners to progress from one activity to another (Khaldi et al., 2023; S. Zhang & Hasim, 2023). Yamani (2021) explored the relationship between gamified learning and UDL principles and suggested incorporating game-informed elements into learning to provide a flexible approach to teaching and learning, rather than as a motivational add-on. Huseinović (2023) and Ismailov and Chiu (2022) noted that learners with different readiness levels and learning styles require sustained oral practice to develop their English speaking skills. To meet this need, research on UDL and gamification has grown, but empirical studies investigating their integration remain limited (Ghai & Tandon, 2022; Luo, 2023). Importantly, most studies on gamification have examined its impact on learner motivation, engagement, and accessibility rather than on language learning outcomes, particularly EFL speaking performance (Rusconi & Squillaci, 2023). Thus, the current study explores whether UDL-based gamified learning is associated with EFL speaking performance and whether differences extend to all four sub-skills of speaking: fluency, grammatical accuracy, vocabulary range, and pronunciation.

Integrating UDL-based Gamified instruction for speaking development

Integrating UDL and gamified instruction provides a coherent basis for designing speaking activities that respond to learner variability while sustaining participation across successive tasks. Griful-Freixenet et al. (2021) conceptualize UDL around multiple means of engagement, representation, and action and expression, treating learner variability as an expected feature of instructional planning. This flexibility matters especially in speaking classes, where learners may differ considerably in proficiency, confidence, learning pace, and willingness to communicate. Gamification addresses a related aspect of instruction by using structured challenges, feedback, goals, and progression to maintain learner involvement across learning activities. Studies in language and educational contexts have associated these features with stronger engagement, participation, and learning outcomes when game elements are closely connected to pedagogical objectives (Al-Dosakee & Ozdamli, 2021; Huseinović, 2023; Laura-De La Cruz et al., 2023; Manzano-León et al., 2021; S. Zhang & Hasim, 2023). The rationale for integration therefore rests on a clear instructional relationship: UDL accommodates variation in how learners access and participate in speaking activities. At the same time, gamified structures organize continued participation and progression through those activities.

Technology offers practical ways to implement this relationship. Digital learning environments can accommodate varied task sequences, feedback formats, and pathways for learners to engage in language activities (Li et al., 2022; Qiao & Zhao, 2023). Mobile and technology-mediated learning can similarly extend oral practice beyond conventional classroom interaction and offer additional ways for learners to participate in and produce spoken language (Cabrera-Solano, 2020; Shortt et al., 2021; Zou et al., 2023). In gamified environments, feedback and visible progress indicators can inform learners about their performance across successive activities and support the regulation of learning effort (W. Zhang et al., 2021). Related EFL evidence links feedback engagement to writing self-efficacy and perceived assessment performance (M. Almayez, 2025), while Orak (2025) shows that the pedagogical use of AI-generated feedback also involves questions of teacher trust, instructional control, and adaptation. Collaborative digital settings add a social dimension through peer interaction, shared tasks, and feedback exchange, all of which support communicative language use (Bizami et al., 2022; Y.-c. Chen, 2022; Ramalingam et al., 2022). The pedagogical value of these features depends on how they are organized, not on the technology itself. Lampropoulos and Kinshuk (2024) draw particular attention to instructional design and teacher competence, including the need to manage competitive and collaborative elements in ways that do not disadvantage particular learners.

Recent intervention research further supports examining this approach in relation to speaking performance. Benlaghrissi and Ouahidi (2024) reported favorable speaking outcomes in mobile-assisted project-based instruction informed by UDL principles, linking flexible instructional design with performance-oriented language learning. In a computer game-based context, Al-Jamili et al. (2024) found improved speaking performance when instruction incorporated structured gameplay and feedback. Hwang et al. (2025) similarly reported improvements in oral performance through multimodal video-dubbing activities, while W. Yan et al. (2024) documented favorable speaking outcomes in immersive learning involving situated interaction and collaborative activity. Although these studies use different instructional models, they provide relevant evidence that flexible participation, structured task engagement, feedback, and technology-mediated interaction can be incorporated into instruction aimed at oral performance. Direct evidence for an explicitly integrated UDL-based gamified framework remains limited, particularly when speaking is assessed both as an overall performance outcome and through fluency, grammatical accuracy, vocabulary range, and pronunciation. Examining these outcomes can clarify whether differences associated with the integrated framework are apparent at the general level of speaking performance and whether the same pattern extends across its component dimensions.

Research gap and hypotheses development

Research on EFL speaking indicates that oral performance comprises related but distinguishable dimensions, including fluency, grammatical accuracy, vocabulary range, and pronunciation, and that development across these dimensions may follow different patterns (Ginther et al., 2010; Tavakoli et al., 2020). Classroom research also points to persistent instructional difficulties in providing sustained oral practice, meaningful interaction, appropriate feedback, and sufficient opportunities for learners with different proficiency levels and communicative readiness to participate (Abdullah et al., 2019; Namaziandost et al., 2019; Saed et al., 2021). Studies comparing different forms of speaking instruction further suggest that changes in instructional organization can produce different patterns of oral performance across speaking outcomes (Asratie et al., 2023; Yenkimaleki & van Heuven, 2019). These findings make speaking instruction design particularly important when the objective extends beyond general participation to measurable performance across specific dimensions of oral production.

UDL and gamification address different aspects of this instructional problem. UDL provides a framework for designing flexible routes to participation and responding to learner variability through multiple forms of engagement, representation, and expression (Rusconi & Squillaci, 2023). Research in language and broader educational settings has associated UDL-informed instruction with accessibility, differentiated participation, and flexible instructional design (Al-Azawei et al., 2016; Lintangsari & Emaliana, 2020; Rao & Meo, 2016). Gamified instruction, in comparison, uses structured goals, task progression, feedback, and game-informed elements to organize learner participation across instructional activities. Studies in language education have reported positive associations with engagement, participation, and speaking performance when these elements are integrated into meaningful language-learning tasks (Huang, 2021; Huseinović, 2023; S. Zhang & Hasim, 2023).

The empirical link between these two lines of research remains underdeveloped. UDL studies in language education have largely focused on accessibility, learner participation, and flexible instructional provision rather than directly assessing speaking performance and its component dimensions (Bray et al., 2023; Ismailov & Chiu, 2022). Research on gamified language learning has focused more on motivation, engagement, participation, and selected language outcomes, with substantial variation in instructional design and the game elements used (Cabrera-Solano, 2020; Khaldi et al., 2023; Luo, 2023). Consequently, evidence from either body of research cannot establish whether combining UDL principles with gamified instructional structures is associated with stronger EFL speaking performance than conventional instruction.

Recent studies involving UDL-informed, game-based, multimodal, and immersive approaches provide relevant evidence that flexible participation, structured task engagement, feedback, and technology-mediated interaction can be incorporated into speaking instruction (Al-Jamili et al., 2024; Benlaghrissi & Ouahidi, 2024; Hwang et al., 2025; W. Yan et al., 2024). However, these studies represent different instructional models and do not directly test an integrated UDL-based gamified framework. Controlled classroom evidence comparing such an integrated framework with conventional speaking instruction remains limited, particularly when speaking is assessed both as an overall performance outcome and through fluency, grammatical accuracy, vocabulary range, and pronunciation. This study addresses this gap through a quasi-experimental comparison of the two instructional conditions. Accordingly, the study addresses the following research question:

RQ1. To what extent does EFL speaking performance differ between learners receiving UDL-based gamified instruction and those receiving conventional speaking instruction, after accounting for baseline performance, in terms of overall speaking performance, fluency, grammatical accuracy, vocabulary range, and pronunciation?

The following central hypothesis and component-level hypotheses were formulated:

H1: Learners receiving UDL-based gamified instruction will demonstrate significantly higher overall EFL speaking performance than learners receiving conventional instruction.

To examine differential effects across the analytic components of speaking performance, the following sub-hypotheses are advanced:

H1a: Learners exposed to UDL-based gamified instruction will demonstrate higher speaking fluency than learners receiving traditional instruction.

H1b: Learners exposed to UDL-based gamified instruction will demonstrate greater grammatical accuracy in speaking than learners receiving traditional instruction.

H1c: Learners exposed to UDL-based gamified instruction will demonstrate a broader range of vocabulary use in speaking than learners receiving traditional instruction.

H1d: Learners exposed to UDL-based gamified instruction will demonstrate clearer and more intelligible pronunciation than learners receiving traditional instruction.

Research Method

Research design

The study employed a quasi-experimental pretest–posttest control-group design. This design fit the instructional context because learners were enrolled in intact classes, and individual random assignment was not feasible. One class received UDL-based gamified instruction, while the other followed conventional instruction. Pre- and post-intervention measures allowed examination of changes in speaking performance over the instructional period, using baseline performance to compare post-intervention outcomes. This design is commonly used in educational settings where existing class arrangements must be retained while comparing instructional approaches (Miller et al., 2020). It also aligns with the study's aim of examining differences in overall speaking performance and its individual components across the two instructional conditions in an authentic classroom setting.

Participants

The study included 42 undergraduate EFL learners enrolled in a compulsory English-speaking course in the preparatory program of a public university in Turkey. Participants ranged in age from 18 to 23 years and included 14 males and 28 females. The experimental group included 8 male and 13 female students, while the control group included 6 male and 15 female students. Eligibility required enrolment in the target speaking course, completion of the prerequisite English courses, and placement at the corresponding proficiency level through the university's institutional placement procedures. All participants followed the same academic track and institutional speaking syllabus. English was a foreign language for all learners, and none reported receiving additional structured English-speaking instruction outside the university curriculum during the study period.

A non-probability convenience sampling strategy was used to recruit two intact classes from the target program. The university's regular administrative procedures had formed the classes before the study began, and they were not created for research purposes. Each class included 21 students. All eligible students in the participating classes were invited to take part voluntarily and provided informed consent before data

collection. Researchers did not randomly assign or transfer individual students between classes for research purposes. One intact class served as the experimental group and received the UDL-based gamified instruction, whereas the other served as the control group and received conventional speaking instruction. The program's existing administrative arrangements determined which instructional condition each intact class received. Both groups received the same speaking syllabus, course content, instructional time, and assessment requirements during the same academic term and were taught by the same instructor.

All 42 participants completed the 14-week instructional period and both the pre-test and post-test assessments, with no attrition. The final analytical sample therefore consisted of 21 learners in each condition. Baseline speaking performance was assessed before the instructional period, and the Results section reports between-group comparisons of pre-test scores. Because only one intact class represented each instructional condition, class membership and instructional condition were confounded, and their respective influences cannot be separated statistically. The findings should consequently be interpreted as differences between the two participating instructional groups rather than as effects attributable exclusively to the instructional framework.

A sensitivity power analysis was conducted for the planned ANCOVA to determine the effect magnitude detectable with the available sample. With $N = 42$, two groups, one covariate, $\alpha = .05$, and statistical power of .80, the analysis indicated that the study could detect an effect of approximately $f = 0.44$ or larger. The available sample was therefore sufficiently sensitive to relatively large between-group effects but had limited power to detect smaller effects. This limitation is particularly relevant when interpreting component-level findings that did not remain statistically significant after multiple-comparisons correction.

Instruments

We assessed speaking performance using an analytic rubric designed for tertiary-level EFL learners. The rubric assessed four speaking areas: fluency, grammatical accuracy, vocabulary range, and pronunciation. These dimensions are consistent with established accounts of speaking proficiency as a multidimensional ability (Ginther et al., 2010; Ounis, 2017; Tavakoli et al., 2020). We rated each dimension from 1 (Very Limited) to 5 (Excellent), with descriptors specifying what performance at each level should look like. Fluency focused on speech continuity, pausing, and temporal control. Grammatical accuracy reflected the appropriate and consistent use of morphosyntactic forms in spontaneous speech. Vocabulary range concerned the diversity, precision, and appropriateness of lexical choices, while pronunciation covered segmental clarity, stress, and overall intelligibility.

The analysis scores were derived directly from the rubric ratings. Each 1–5 rating was multiplied by five, producing a score from 5 to 25 for each speaking dimension. The four component scores were then summed, giving an overall speaking score ranging from 20 to 100. The rescaling expressed the rubric ratings on the reporting scale adopted for the component and overall speaking scores; as a linear transformation, it did not alter relative differences among participants or statistical comparisons. The overall score represents the sum of the four rescaled component scores and is not a percentage. Retaining the four component scores also allowed examination of differences in fluency, grammatical accuracy, vocabulary range, and pronunciation separately from overall speaking performance. This distinction was important because development may vary across individual dimensions of oral performance (Namaziandost, 2019; Tuan, 2012). Appendix (Table III) provides the full rubric and its performance descriptors.

Rubric validation

Given the use of analytic scoring and human judgment, inter-rater reliability was examined for the speaking assessment. Two raters with experience in EFL speaking assessment independently evaluated all recorded performances using the same analytic rubric. Before formal scoring, the raters reviewed the four assessment dimensions, fluency, grammatical accuracy, vocabulary range, and pronunciation, and discussed the performance descriptors associated with each scoring level. A preliminary calibration exercise used sample recordings to establish a shared interpretation of the scoring criteria before the main ratings. The calibration recordings were not included in the reliability analysis. During formal scoring, the raters evaluated the recordings independently and were blinded to participants' instructional condition and testing occasion.

Inter-rater reliability was estimated using a two-way random-effects Intraclass Correlation Coefficient (ICC) based on absolute agreement. Because the statistical analyses used the mean of the two independent ratings, we used average-measures ICCs to assess the reliability of the final scores. ICC estimates ranged from .81 to .88 across the four speaking components, with an ICC of .88 for the overall speaking score, indicating a high degree of agreement between the two raters. The mean of the two independent ratings was retained as the final score for each speaking component, and the resulting component scores were summed to derive the overall speaking-performance score. The averaging procedure retained differences between raters rather than resolving them through post-rating discussion or consensus scoring.

Evidence supporting the content validity of the speaking rubric was established before its use in the main study. The initial rubric was developed with reference to established conceptualizations of oral proficiency and analytic approaches to EFL speaking assessment (Ginther et al., 2010; Ounis, 2017; Tavakoli et al., 2020). It operationalized speaking performance through four dimensions—fluency, grammatical accuracy, vocabulary range, and pronunciation—with five performance levels specified for each dimension. The initial version was reviewed by two

specialists in applied linguistics and EFL assessment. The reviewers examined the relevance of the four dimensions to the targeted speaking construct, the correspondence between the descriptors and their respective dimensions, the clarity and progression of the five performance levels, and the suitability of the rubric for undergraduate EFL learners. Their comments were used to refine the wording of several descriptors and improve distinctions between adjacent performance levels. The revised rubric was then piloted using recorded speaking performances from learners not included in the main study. The pilot scoring was used to examine whether the descriptors could be applied consistently and whether any scoring categories required further clarification. Minor wording adjustments were made before the final version was adopted. The resulting rubric retained the four theoretically grounded dimensions and five-level analytic structure, with the complete scoring descriptors provided in Appendix (Table III). This development and review process provided content-based evidence that the rubric was appropriate for assessing the study's targeted speaking performances.

Intervention details

The instructional intervention was implemented across a fourteen-week academic term, with two 45-minute speaking sessions per week integrated into the regular course schedule. Both the experimental and control groups followed the same institutional syllabus and addressed identical communicative themes prescribed for the speaking course. Instructional time, course content, and assessment requirements were equivalent across conditions.

The same instructor taught both groups to control for teacher-related variation. The instructor did not score the speaking assessments; raters evaluated them independently and were blinded to group membership. All sessions took place during scheduled class hours in standard instructional classrooms.

The experimental condition differed from the control condition in how speaking practice was organized, not in content coverage or time allocation. Speaking activities in the experimental group followed a UDL-based gamified framework designed to accommodate learner variability in participation, pacing, and oral production. Tasks offered multiple modes of engagement and expression, allowing learners to choose participation formats aligned with their linguistic readiness while working toward shared communicative objectives.

Specifically, the UDL-based gamified framework integrated the core UDL principles of multiple means of engagement, action, and expression into the organization of speaking tasks. Learners were offered flexible participation pathways, differentiated entry points, and varied modes of oral contribution, allowing them to engage with speaking tasks in ways aligned with their linguistic readiness while pursuing shared communicative goals. Gamified elements such as task sequencing, visual progress indicators, and staged advancement structured sustained speaking practice and supported continuity and feedback alignment, rather than introducing competition or reward-driven play. For example, in a role-based speaking task, learners selected from alternative participation roles (e.g., primary speaker, discussion moderator, or responder), progressed through sequenced communicative challenges of increasing complexity, and received criterion-referenced feedback aligned with fluency, grammatical accuracy, vocabulary range, and pronunciation upon completion of each task cycle.

The speaking practice was organized in sessions with increasingly complex communicative tasks. The learner practiced role plays, problem-solving tasks for specific situations, and discussions. The complexity of the tasks and roles, as well as the participation format, could be adjusted to provide different entry points into the practice that were equivalent in terms of the goals of communicative competence. The gamified elements organized the progression and the continuity of speaking practice. Learners earned points for completing speaking tasks, could check their progress using progress bars, and proceeded to the next speaking challenge after completing one. These elements organized the speaking practice rather than fostering competition.

Instructor involvement focused on facilitation and monitoring of interaction. The instructor provided guidance at task boundaries, allowing learners to maintain extended speaking turns during activities. Feedback addressed recurring patterns observed during speaking practice and was aligned with the analytic speaking rubric. Periodic reflection was incorporated, during which learners reviewed speaking outcomes against task goals and identified areas needing further attention.

The control group completed the same speaking topics using conventional instructional procedures. Speaking activities followed textbook-guided formats and routine pair-based or whole-class interaction without differentiated task pathways or gamified organization. Participation patterns and task pacing were uniform across learners, and the instructor provided feedback in accordance with standard classroom practice.

To maintain comparability between conditions, we assigned no additional speaking practice outside scheduled class sessions, and we asked learners not to engage in structured external speaking activities related to the course during the intervention period. Both groups completed identical pre-test and post-test speaking assessments, which raters blinded to group membership evaluated using the same analytic rubric; the only systematic difference between conditions was how we organized speaking practice during classroom sessions.

We used a structured implementation checklist to monitor implementation fidelity throughout the 14-week intervention period. The checklist detailed the key instructional features of the teaching framework that were implemented in experimental sessions. The principal features were (1) flexible participation pathways; (2) a range of possible entry points; (3) gamified task sequencing; (4) visual progress indicators; (5) staged advancement; (6) point accumulation; and (7) criterion-referenced feedback for student performance. The instructor monitored implementation by completing a checklist after each session and keeping brief teaching logs recording what was delivered and any deviations from

the planned implementation. The control group also received conventional instruction using a textbook as an aid. The research team reviewed implementation records periodically throughout the implementation phase to monitor consistency and deviations between teaching sessions in both the experimental and control groups.

Data collection

Data were collected during the first term of the 2025–2026 academic year, following a pre-intervention, instructional, and post-intervention sequence. Before the instructional period began, all participants completed an individual speaking pre-test to establish their baseline performance in fluency, grammatical accuracy, vocabulary range, pronunciation, and overall speaking performance. The assessment was conducted in a controlled setting using speaking tasks aligned with the course content and communicative demands. Participants were given the same task instructions and assessment conditions, and each performance was audio-recorded for subsequent rating.

Following the pre-test, the two classes completed the 14-week instructional period. The experimental group received speaking instruction through a gamified, Universal Design for Learning-based framework, whereas the control group received conventional speaking instruction. Both groups studied the same course content and received the same amount of instructional time. The study did not assign additional speaking practice outside the scheduled sessions.

At the end of the instructional period, participants completed an individual speaking post-test. We used parallel prompts at the two measurement points to minimize familiarity with the assessment material while maintaining comparable task structure, duration, and communicative demands. We followed the same administration procedures for the post-test and audio-recorded all performances. All 42 participants completed both assessments; consequently, no cases were excluded because of missing pre-test or post-test data.

Before scoring, we anonymized the recordings and assigned identification codes that did not reveal instructional condition or assessment occasion. Two independent raters evaluated the recordings using the analytic speaking-performance rubric described in the preceding section. The raters were blinded to both group membership and whether a recording had been obtained at pre-test or post-test. The resulting ratings constituted the dataset used to examine between-group differences in overall speaking performance and the four component scores.

Ethical considerations

The Fırat University Social and Human Sciences Research Ethics Committee approved the current study (approval code: 03.06.2025-35147 dated 29 May 2025). All participants received a clear description of the objectives of the study as well as the procedures that were followed to gather and use the data. Participation in the study was voluntary. All participants signed informed consent forms before participating. However, participants were informed that their decision to participate or not would negatively affect their academic evaluation, course grades, or relationship with the program. All speech samples were analyzed without any identifying information about the participants. The data, including the audio recordings and assessment results, were stored securely, and only the researchers had access. All the procedures within the scope of the current study were conducted in line with the ethical rules for involving human participants in research within the Fırat University Social and Human Sciences.

Data analysis

The data analysis focused on post-intervention differences in speaking performance between the control and experimental groups while accounting for baseline variation. All speaking scores were screened before analysis for completeness and consistency across measurement points. The dataset contained no missing values, as all participants completed the pre-test and post-test assessments for each speaking component and the overall score.

We examined pre-test differences between the two groups using independent-samples t-tests for fluency, grammatical accuracy, vocabulary range, accent, and total speaking ability. These analyses described and tested for between-group differences at pre-test and therefore did not test for statistical equivalence. We examined post-test differences between the two groups using separate Analyses of Covariance (ANCOVA) for each speaking ability component and the overall score. In each analysis, we used the post-test score as the dependent variable, the instructional group as the fixed factor, and the corresponding pre-test score as the covariate. This allowed us to compare post-test performance while controlling for pre-test differences in the relevant measure.

Before estimation of the ANCOVA models, the relevant assumptions were examined for each outcome. Linearity between the covariate and the corresponding post-test score was assessed through graphical inspection of their relationship and model residuals. Homogeneity of regression slopes was examined by testing the interaction between instructional condition and the corresponding pre-test covariate. Residual distributions were inspected using Q–Q plots to assess substantial departures from normality, and homogeneity of error variance was examined using Levene's test. No substantial violations were identified that precluded the use of ANCOVA. For each model, adjusted post-test means were estimated for both groups, and between-group differences were evaluated using F statistics with their associated degrees of freedom and p values. Partial eta squared $(\eta^2 p)$ was reported as an index of effect size.

Overall speaking performance was the study's overarching outcome, while fluency, grammatical accuracy, vocabulary range, and pronunciation addressed the component-level hypotheses. Because separate ANCOVA models were conducted for the four conceptually related speaking components, the Holm procedure was applied to their p-values to control for the family wise Type I error rate. We reported both unadjusted and Holm-adjusted *p*-values for the component-level analyses, while evaluating overall speaking performance separately. Statistical significance was evaluated at an alpha level of .05.

Findings

A pre-test comparison was conducted to examine whether statistically significant differences in speaking performance were present between the two groups before the intervention. Table I presents the descriptive statistics and independent-samples t-test results for each speaking component and the overall score. No statistically significant between-group differences were detected for fluency (t = −0.19, p = .847), grammatical accuracy (t = −0.39, p = .701), vocabulary range (t = −0.71, p = .482), pronunciation (t = −0.19, p = .850), or overall speaking performance (t = −0.59, p = .558). The descriptive statistics also showed relatively small differences between the group means across the measured outcomes. These findings indicate that the pre-test comparisons did not detect statistically significant differences in speaking performance between the two groups. Given the sample size and the non-randomized design, the absence of statistically significant differences should not be interpreted as evidence of baseline equivalence. The corresponding pre-test scores were subsequently included as covariates in the ANCOVA models to account for baseline variation when comparing post-test performance.

Table I
Pre-test descriptive statistics and between-group comparisons

Criterion	Control		Experimental		*t*-value	*p*-value
	Mean	SD	Mean	SD		
Fluency	13.81	7.89	14.29	7.95	−0.19	.847
Accuracy	14.52	7.89	15.48	8.05	−0.39	.701
Vocabulary	14.05	7.00	15.48	6.50	−0.71	.482
Pronunciation	14.05	8.16	14.52	8.05	−0.19	.850
Total Score	56.43	16.74	59.76	19.72	−0.59	.558

Table II presents post-test outcomes for the control and experimental groups, including raw means and ANCOVA-derived adjusted means. The adjusted means represent post-test performance after accounting for the corresponding pre-test scores. For each speaking component and the overall score, the table reports the F statistic, p-value, and partial η^2.

Table II
Post-test adjusted means and ANCOVA results for EFL speaking performance

Criterion	Control		Experimental		Control	Experimental	$F(1, 39)$	p	Holm-adjusted p	Partial η^2
	Mean	SD	Mean	SD	Adjusted Mean	Adjusted Mean				
Fluency	12.86	6.63	17.62	7.35	12.92	17.55	4.95	.032	.064	0.11
Accuracy	10.48	5.22	20.24	6.02	10.55	20.17	31.13	< .001	< .004	0.45
Vocabulary	13.33	6.77	18.10	6.61	13.51	17.92	4.68	.037	.064	0.11
Pronunciation	14.76	7.33	20.24	6.22	14.77	20.23	6.61	.014	.042	0.15
Total Score	51.43	15.01	76.19	17.24	52.10	75.52	26.88	< .001	—	0.41

After adjusting for corresponding pre-test scores, the experimental group had higher adjusted post-test means than the control group across all four speaking components. For fluency, adjusted means were 17.55 for the experimental group and 12.92 for the control group, $F(1, 39)$ = 4.95, p = .032, partial η^2 = .11. Following Holm correction for multiple component-level comparisons, this difference did not remain statistically significant (adjusted p = .064). The adjusted means were 20.17 for the experimental group and 10.55 for the control group, $F(1, 39)$ = 31.13, p < .001, partial η^2 = .45. The difference remained statistically significant after Holm correction (adjusted p < .004).

For vocabulary range, the experimental group had a higher adjusted mean (17.92) than the control group (13.51), $F(1, 39)$ = 4.68, p = .037, partial η^2 = .11. This difference did not remain statistically significant after Holm correction (adjusted p = .064). For pronunciation, the adjusted means were 20.23 for the experimental group and 14.77 for the control group, $F(1, 39)$ = 6.61, p = .014, partial η^2 = .15. The difference remained statistically significant after Holm correction (adjusted p = .042).

For overall speaking performance, the adjusted post-test mean was 75.52 for the experimental group and 52.10 for the control group, $F(1, 39)$ = 26.88, p < .001, partial η^2 = .41. The primary outcome thus showed a statistically significant between-group difference after adjustment for baseline performance. Among the component-level outcomes, grammatical accuracy and pronunciation remained statistically significant

after correction for multiple testing, whereas fluency and vocabulary range did not. The largest partial η^2 values were observed for grammatical accuracy (.45) and overall speaking performance (.41), followed by pronunciation (.15), fluency (.11), and vocabulary range (.11).

Figure 1 presents the pre-test and post-test mean scores for both groups. The experimental group recorded higher post-test scores across all four speaking components and overall speaking performance, whereas the control group's change was less consistent. The figure visually represents the descriptive patterns reported in Tables I and II.

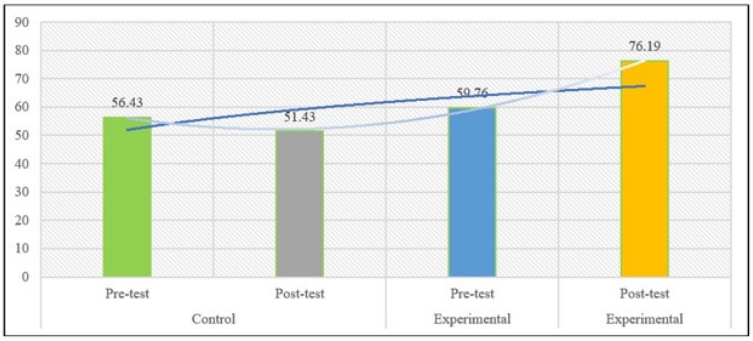


Fig. 1 *Pre-test and post-test mean scores for the control and experimental groups*

Discussion

This study examined the effects of a UDL-based gamified environment on speakers of other languages in EFL speaking situations after a lesson. The findings revealed significant differences between the two groups, with the experimental group achieving higher adjusted post-test mean scores than the control group even after adjusting for pre-test means. Overall speaking ability differed significantly between the two groups, with a large effect size (partial η^2 = .41). However, the four component measures told a more complex story. The experimental group showed higher adjusted means for four components: Fluency, grammatical accuracy, vocabulary range, and pronunciation. However, only two components (grammatical accuracy and pronunciation) remained significant after applying Holm's method to control family-wise error rates. Although fluency and vocabulary range were significant in analyses that did not control for family-wise error rates, the evidence did not support including them as separate components in the experimental group's scores. In summary, the strongest support for the finding that the experimental group of students demonstrated improved speaking ability as a result of learning in the UDL-based gamified environment was for grammatical accuracy and pronunciation.

For grammatical accuracy, the largest effect size was observed (partial η^2 = .45), and even after applying the multiple-testing correction, the result remained statistically significant. The experimental design for communicative practice and criterion-referenced feedback plausibly explains this strong result for grammatical accuracy. Learners completed a long speaking task and later received feedback on recurring aspects of their performance, allowing attention to grammatical form without continually breaking into meaning-focused components of speaking. This delayed, criterion-referenced feedback allows learners to notice and consolidate form–meaning relationships (Ghahderijani et al., 2021; Navidinia et al., 2019). Repeated task engagement may also give learners opportunities to reconsider grammatical choices across successive speaking attempts, a pattern reported in previous classroom research (Afifah & Devana, 2020). Similar associations between structured feedback and grammatical development have been reported in technology-supported instructional settings (Mulyadi et al., 2021; Qiao & Zhao, 2023). In the present study, these instructional features reasonably account for the large between-group difference in grammatical accuracy, although the available data cannot establish the mechanism. We did not directly measure feedback uptake, noticing, or changes in grammatical production across successive tasks, leaving the specific processes underlying the observed accuracy difference open to further investigation.

Pronunciation also showed a statistically reliable between-group difference after Holm correction, although its effect size was smaller than that observed for grammatical accuracy. A plausible explanation concerns the communicative demands built into the experimental activities. Role-based interaction, problem-solving tasks, and extended discussions required learners to make themselves comprehensible to their interlocutors to sustain interaction and achieve shared communicative goals. This communicative orientation may have encouraged greater attention to intelligibility during oral production. Previous research has associated communicatively situated pronunciation practice with intelligibility development, particularly when learners negotiate meaning and respond to interlocutors during interaction (Yenkimaleki & van Heuven, 2019). Related studies in technology-supported speaking contexts link pronunciation development to feedback and repeated communicative practice focused on intelligibility rather than isolated phonetic production (Bashori et al., 2022; Sun, 2023; Tai, 2022; Tai & Chen, 2022). This result aligns with this line of research and suggests that embedding pronunciation demands within purposeful communication may relate to the observed performance difference. This interpretation remains provisional, since the study did not directly assess phonological monitoring, interlocutor responses, or learners' pronunciation adjustments during classroom interaction. The findings thus establish a difference in pronunciation performance between

the two instructional conditions, rather than identifying the process through which that difference developed.

Fluency data showed a positive change for the experimental group, with the adjusted post-test mean greater than the pre-test mean. The unadjusted ANCOVA revealed a significant main effect for group, with the experimental group performing significantly better than the control group. However, after applying the Holm correction for multiple testing, the difference between the groups was no longer significant at the 0.05 level. Thus, whilst the data show a positive change in the fluency of the experimental group and indeed show that this group outperformed the control group as found with the unadjusted test, there is no strong evidence to suggest that there is a reliable component-level advantage for this group after taking into account the effects of testing on the different measures of oral production. Sustained oral interaction in the experimental activities could explain the positive change in the experimental group's fluency. That is, all the experimental activities involved learners in repeated meaning-focused speaking and therefore provided more opportunities for extended production than the control activities. Previous research shows that extended oral production facilitates fluency development (De Jong, 2018; Derakhshan et al., 2016; Tavakoli et al., 2020). Moreover, the extended oral interaction provided by the experimental activities may have helped learners engage in more sustained speech and, consequently, maintain a high level of speech continuity throughout communicative events and turns (Buckingham & Alpaslan, 2017; Tsou, 2005). However, the study did not assess speaking time, speaking rate, pausing, or learners' oral production. Thus, although the data provide some evidence of positive change in the experimental group's fluency, this should be viewed with caution, as it is merely descriptive and does not provide strong evidence of a component-level advantage over the control group.

Vocabulary range followed a comparable statistical pattern. The experimental group recorded a higher adjusted post-test mean, but the difference was no longer statistically significant after Holm correction. The observed direction may relate to the communicative demands of the experimental activities, although this interpretation remains tentative. Role-based interaction, situational problem solving, and guided discussion required learners to select lexical resources in response to changing contexts, interlocutors, and communicative purposes. Previous research has associated task-based interaction and contextualized language use with broader lexical deployment in speaking (Janebi Enayat & Derakhshan, 2021; Kilic, 2019) while collaborative and project-oriented speaking activities have been linked to repeated use of lexical resources across varied communicative contexts (Benlaghrissi & Ouahidi, 2024; Cabrera-Solano, 2020; Darmuki et al., 2018; Huang, 2021). These instructional characteristics plausibly account for the higher adjusted vocabulary mean observed in the experimental group. Still, the present data cannot identify them as the source of that difference. Lexical production during classroom interaction was not analyzed, and the corrected component-level comparison was not statistically significant. The vocabulary result should thus be regarded as a favorable descriptive pattern that merits further investigation rather than evidence of a reliable vocabulary advantage associated with the experimental condition.

A broader perspective on speaking performance is also important. So far, much research has examined gamified EFL instruction, linking gamification to greater participation, higher motivation to use speech, and decreased anxiety when using speech in class. However, previous studies have provided less consistent evidence of increased assessed speaking performance (Huang, 2021; Huseinović, 2023; Luo, 2023). The present study examines the broader effects of gamified speech practice on students' speaking performance. Unlike many studies that use game elements, such as rewards, badges, and leaderboards, as motivational tools added to a more traditional classroom setting, gamified elements are an integral part of a pedagogically grounded speech practice framework. The framework significantly improved students' overall speaking performance. Gamified elements are integral to how language practice is organized, rather than added motivators in a more conventional classroom.

Furthermore, the component-wise analysis yielded significant results for two of the four speaking components: grammatical accuracy and lexical orchestration (pronunciation). For the other two components of speaking, namely fluency and vocabulary range, the gamified treatment did not show a significant positive effect. This overall difference, despite very different results across the four components, suggests that the positive association between gamified language practice and enhanced speaking ability is not fixed or generalizable across all four components. Still, it is specific and highly variable, depending on each component's linguistic demands and the performance demands required to use it effectively.

Another basis for interpreting the speaking results is UDL integration. While many studies have examined UDL, they primarily focus on accessibility, participation, and learners' perceptions (Al-Azawei et al., 2016; Bray et al., 2023; Ismailov & Chiu, 2022; Rusconi & Squillaci, 2023). However, in the present study, within the framework of the speaking activities, the principles of UDL were used to design flexible participation, allow different starting points, and include different ways of oral contribution to the shared communicative goals. In contrast, the gamified elements structured participation and made learners' progress in individual activities visible. An important aspect of combining these two elements is that they address different instructional problems. While UDL addresses the variety of ways learners can access and participate in speaking activities, the gamified organization provides structure and progression within individual activities. Therefore, the significant difference in speaking results between the two conditions suggests that combining these two elements may be a promising approach for performance-oriented EFL instruction. It extends UDL beyond accessibility and suggests that gamification offers more than motivation.

The study design constrains the interpretation of the findings. In the 14-week study, the differences between the two instructional conditions (i.e., flexible participation, differentiated task entry, structured progression, repeated interaction, feedback, and gamified features) cannot be isolated and attributed to individual features. Instead, these features were embedded within a single instructional condition, and it is not possible to determine the effects of individual components. Furthermore, the study did not examine classroom-environment factors that could account for participants' speaking performance, such as interaction with the teacher and other students, engagement with feedback, uptake of feedback,

and use of strategies. Hence, any proposed accounts of this study's data should be viewed as theoretically grounded interpretations rather than empirically established mechanisms. Subsequent research should aim to investigate these processes in more detail through methods such as classroom observation, interactional and discourse analysis, and mediation studies. Comparative studies that systematically vary individual features of the UDL-based gamified framework could also determine which instructional features best support which aspects of speaking performance.

Theoretical and practical implications

The findings from this study could extend the theoretical basis for integrating UDL and gamification in EFL speaking classes. The significant difference in overall speaking performance provides evidence that a UDL-based gamification framework can influence assessed oral performance. However, this study does not directly measure outcomes such as participants' participation levels, access to learning opportunities, motivation to learn, and related outcomes. Importantly, the component scores were also significant, especially for grammatical accuracy and pronunciation, even after correcting for the family-wise error rate for multiple comparisons. However, fluency and vocabulary range did not receive similar support. Thus, while speaking development is often treated as a single entity that can be enhanced through a particular pedagogical approach, this study shows that different oral communicative abilities develop in different ways, even when learners are exposed to the same study condition. In other words, the study's primary theoretical contribution is linking UDL-based gamification to performance-based measures of speaking ability while distinguishing overall speaking performance from its key components.

For classroom practice, the findings support an approach that embeds flexibility and gamified progression within speaking instruction rather than adding them as peripheral features. Teachers may structure speaking tasks around different participation pathways and entry points while maintaining common communicative goals and assessment criteria. The stronger findings for grammatical accuracy and pronunciation also support giving these dimensions deliberate attention within communicative tasks, particularly through criterion-referenced feedback and activities that require comprehensible oral production. Greater caution is needed when interpreting the findings for fluency and vocabulary range because correcting for multiple comparisons made their component-level differences no longer statistically significant. The practical implication is thus not that UDL-based gamification produces equivalent improvement across speaking dimensions, but that it provides an instructional structure within which specific dimensions can be addressed through appropriately targeted speaking activities.

Limitations and directions for future research

Several limitations should be considered when interpreting the findings. First, the quasi-experimental design involved two intact classes, with one class assigned to each instructional condition. Therefore, we could not analytically separate class membership from instructional condition, and unmeasured differences between the two classes may have contributed to the observed post-intervention differences. Although both groups followed the same syllabus and instructional schedule, were taught by the same instructor, and the analysis accounted for baseline speaking performance, these controls cannot rule out possible class-level influences. Future research should include multiple classes within each instructional condition, preferably with random assignment at the class level, to distinguish instructional differences from between-class variation.

Using the same instructor for both groups presents a related limitation. Although this reduced differences attributable to teacher identity, it does not rule out possible differences in the instructor's enthusiasm or attention when delivering the two forms of instruction. The novelty of the UDL-based gamified framework may have influenced its delivery in ways that were not measured in the present study. Reliance on a single instructor also means that the findings reflect one teacher's implementation of the framework and may not generalize to instructors with different teaching styles, experience, or familiarity with UDL and gamification. Future research could involve multiple instructors or independent classroom observations to examine these possibilities.

Second, the instructional period allowed us to examine speaking development during the intervention. Still, it remains unclear whether the observed differences would persist over time or transfer to other speaking contexts. Future studies with delayed post-tests could examine whether these differences persist over time. Third, the study assessed speaking performance through analytic scores without complementary evidence of learners' strategies, classroom interaction, or engagement with instructional activities. Classroom observations, discourse-based evidence, or learner reflections could provide further insight into the processes associated with differences in speaking performance.

Fourth, the study was conducted at a single institution with a relatively small sample, which limits the extent to which the findings can be generalized to other EFL populations and educational settings. Replication across institutions and learner populations is needed to determine whether similar patterns are found in other contexts. Finally, the study did not examine whether learner characteristics such as initial proficiency, previous experience with gamified instruction, or self-regulation were related to different responses to the instructional framework. Future studies with larger samples could examine these learner-level differences.

Conclusion

This study provides evidence that UDL-based gamified instruction can serve as a viable framework for EFL speaking pedagogy. Learners exposed to the framework demonstrated stronger overall speaking performance than those receiving conventional instruction, after accounting for baseline performance. The component-level results add an important qualification to this finding. The clearest differences were in grammatical accuracy and pronunciation, while evidence for fluency and vocabulary range was less conclusive after adjusting for multiple comparisons. The results show that speaking performance is not a single entity and therefore should not be measured in the same way when evaluating the effectiveness of different language-learning interventions. The study found that a pedagogical framework incorporating UDL and gamification effectively improved learners' overall oral ability, with notable gains in some aspects of oral performance. Importantly, however, the effectiveness of individual features of the framework varied, as did the extent to which different aspects of oral ability were affected. This study's main contribution is to move the discussion of UDL and gamification beyond accessibility and motivation to consider their relevance to language performance. Above all, the findings provide a useful basis for exploring how different instructional designs for developing speaking ability relate to specific aspects of speaking performance. In particular, the use of learner variability, purposeful oral participation, and a structured organizational framework is a credible approach in EFL pedagogy.

Funding

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2026-2905-09".

References

Abdullah, M. Y., Hussin, S., & Ismail, K. (2019). Implementation of flipped classroom model and its effectiveness on English speaking performance. *International Journal of Emerging Technologies in Learning*, *14*(09), 130. https://doi.org/10.3991/ijet.v14i09.10348

Afifah, N., & Devana, T. (2020). Speaking skill through task based learning in English foreign language classroom. *Journal of English Language Teaching*, *7*(2), 135–144. https://doi.org/10.33394/jo-elt.v7i2.3109

Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2016). Universal Design for Learning (UDL): A content analysis of peer reviewed journals from 2012 to 2015. *Journal of the Scholarship of Teaching and Learning*, *16*(3), 39–56. https://doi.org/10.14434/josotl.v16i3.19295

Al-Dosakee, K., & Ozdamli, F. (2021). Gamification in teaching and learning languages: A systematic literature review. *Revista Romaneasca Pentru Educatie Multidimensionala*, *13*(2), 559–577. https://doi.org/10.18662/rrem/13.2/436

Al-Jamili, O., Aziz, M., Mohammed, F., Almogahed, A., & Alawadhi, A. (2024). Evaluating the efficacy of computer games-based learning intervention in enhancing English speaking proficiency. *Heliyon*, *10*(16), e36440. https://doi.org/10.1016/j.heliyon.2024.e36440

Al-khresheh, M. H. (2025). Exploring the role of motivation in self-regulated digital game-based vocabulary learning among Saudi learners: A structural equation modelling approach to attention, relevance, confidence and satisfaction. *European Journal of Education*, *60*(4). https://doi.org/10.1111/ejed.70335

Al-khresheh, M. H. (2025). Talking with machines: Experimental evidence on the effects of AI-supported speaking practice on fluency and anxiety. *Applied Linguistics: Research, Measurement and Practice*, *1*(1), 24–44. https://doi.org/10.65334/pmqzf567

Al-khresheh, M. H., & Alruwaili, S. F. (2026). Examining ChatGPT-supported speaking practice among Saudi EFL students: A quasi-experimental study. *Humanities and Social Sciences Communications*, *13*(1), 1433. https://doi.org/10.1038/s41598-024-66047-4

Almayez, M. (2025). Modelling the relationships among feedback engagement, writing self-efficacy, and assessment performance among EFL learners: A structural equation modelling approach. *Applied Linguistics: Research, Measurement and Practice*, *1*(1), 45–73. https://doi.org/10.65334/p7n0vj47

Almayez, M. A., Al-khresheh, M. H., AL-Qadri, A. H., Alkhateeb, I. A., & Alomaim, T. I. M. (2025). Motivation and English self-efficacy in online learning applications among Saudi EFL learners: Exploring the mediating role of self-regulated learning strategies. *Acta Psychologica*, *254*, 104796. https://doi.org/10.1016/j.actpsy.2025.104796

Alrasheedi, S. (2020). Investigation of factors Influencing speaking performance of Saudi EFL learners. *Arab World English Journal*, *11*(4), 66–77. https://doi.org/10.24093/awej/vol11no4.5

Asratie, M. G., Wale, B. D., & Aylet, Y. T. (2023). Effects of using educational technology tools to enhance EFL students speaking performance. *Education and Information Technologies*, *28*(8), 10031–10051. https://doi.org/10.1007/s10639-022-11562-y

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners vocabulary, speaking anxiety, and language enjoyment. *System*, *99*, 102496. https://doi.org/10.1016/j.system.2021.102496

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2022). Look, I can speak correctly: Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, *37*(5-6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230

Benlaghrissi, H., & Ouahidi, L. M. (2024). The impact of mobile-assisted project-based learning on developing EFL students speaking skills. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00303-y

Bizami, N. A., Tasir, Z., & Kew, S. N. (2022). Innovative pedagogical principles and technological tools capabilities for immersive blended learning: A systematic literature review. *Education and Information Technologies*, *28*(2), 1373–1425. https://doi.org/10.1007/s10639-022-11243-w

Bourzeg, M. (2025). Emotion regulation in English language learning: The roles of cognitive reappraisal and expressive suppression among Algerian EFL students. *Applied Linguistics: Research, Measurement and Practice*, *1*(1), 95–113. https://doi.org/10.65334/kkbhp157

Bray, A., Devitt, A., Banks, J., Sanchez Fuentes, S., Sandoval, M., Riviou, K., Byrne, D., Flood, M., Reale, J., & Terrenzio, S. (2023). What next for universal design for learning? A systematic literature review of technology in implementations at second level. *British Journal of Educational Technology*, *55*(1), 113–138. https://doi.org/10.1111/bjet.13328

Buckingham, L., & Alpaslan, R. S. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System*, *65*, 25–37. https://doi.org/10.1016/j.system.2016.12.016

Cabrera-Solano, P. (2020). The use of digital portfolios to enhance English as a foreign language speaking skills in higher education. *International Journal of Emerging Technologies in Learning*, *15*(24), 159. https://doi.org/10.3991/ijet.v15i24.15103

Chen, R. H. (2022). Effects of deliberate practice on blended learning sustainability: A community of inquiry perspective. *Sustainability*, *14*(3), 1785. https://doi.org/10.3390/su14031785

Chen, Y.-c. (2022). Effects of technology-enhanced language learning on reducing EFL learners public speaking anxiety. *Computer Assisted Language Learning*, *37*(4), 789–813. https://doi.org/10.1080/09588221.2022.2055083

Darmuki, A., Andayani, A., Nurkamto, J., & Saddhono, K. (2018). The development and evaluation of speaking learning model by cooperative approach. *International Journal of Instruction*, *11*(2), 115–128. https://doi.org/10.12973/iji.2018.1129a

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237–254. https://doi.org/10.1080/15434303.2018.1477780

Derakhshan, A., Khalili, A. N., & Beheshti, F. (2016). Developing EFL learners speaking ability, accuracy and fluency. *English Language and Literature Studies*, *6*(2), 177. https://doi.org/10.5539/ells.v6n2p177

Dobinson, K. L., & Dockrell, J. E. (2021). Universal strategies for the improvement of expressive language skills in the primary classroom: A systematic review. *First Language*, *41*(5), 527–554. https://doi.org/10.1177/0142723721989471

Erdiana, N., Bahri, S., & Akhmal, C. N. (2019). Male vs. female EFL students: Who is better in speaking skill? *Studies in English Language and Education*, *6*(1), 131–140. https://doi.org/10.24815/siele.v6i1.13024

Ghahderijani, B. H., Namaziandost, E., Tavakoli, M., Kumar, T., & Magizov, R. (2021). The comparative effect of Group Dynamic Assessment (GDA) and Computerized Dynamic Assessment (C-DA) on Iranian upper-intermediate EFL learners speaking Complexity, Accuracy, and Fluency (CAF). *Language Testing in Asia*, *11*(1). https://doi.org/10.1186/s40468-021-00144-3

Ghai, A., & Tandon, U. (2022). Integrating gamification and instructional design to enhance usability of online learning. *Education and Information Technologies*, *28*(2), 2187–2206. https://doi.org/10.1007/s10639-022-11202-5

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379–399. https://doi.org/10.1177/0265532210364407

Griful-Freixenet, J., Struyven, K., & Vantieghem, W. (2021). Exploring pre-service teachers beliefs and practices about two inclusive frameworks: Universal design for learning and differentiated instruction. *Teaching and Teacher Education*, *107*, 103503. https://doi.org/10.1016/j.tate.2021.103503

Hanifa, R. (2018). Factors generating anxiety when learning EFL speaking skills. *Studies in English Language and Education*, *5*(2), 230–239. https://doi.org/10.24815/siele.v5i2.10932

Huang, H.-W. (2021). Effects of smartphone-based collaborative vlog projects on EFL learners speaking performance and learning engagement. *Australasian Journal of Educational Technology*, *37*(6), 18–40. https://doi.org/10.14742/ajet.6623

Huseinović, L. (2023). The Effects of gamification on student motivation and achievement in learning English as a foreign language in higher education. *MAP Education and Humanities*, *4*(1), 10–36. https://doi.org/10.53880/2744-2373.2023.4.10

Hwang, G.-J., Fathi, J., & Rahimi, M. (2025). Fostering learners speaking skills and flow experience with video-dubbing tasks: A flow theory perspective. *Journal of Computer Assisted Learning*, *41*(2), e13120. https://doi.org/10.1111/jcal.13120

Ismailov, M., & Chiu, T. K. F. (2022). Catering to inclusion and diversity with universal design for learning in asynchronous online education: A self-determination theory perspective. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.819884

Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, *99*, 102521. https://doi.org/10.1016/j.system.2021.102521

Khaldi, A., Bouzidi, R., & Nader, F. (2023). Gamification of E-learning in higher education: A systematic literature review. *Smart Learning Environments*, *10*(1), 1–31. https://doi.org/10.1186/s40561-023-00227-z

Kilic, M. (2019). Vocabulary knowledge as a predictor of performance in writing and speaking: A case of Turkish EFL learners. *PASAA*, *57*(1), 133–164. https://doi.org/10.58837/chula.pasaa.57.1.6

Lampropoulos, G., & Kinshuk. (2024). Virtual reality and gamification in education: A systematic review. *Educational Technology Research and Development*, *72*(3), 1691–1785. https://doi.org/10.1007/s11423-024-10351-3

Laura-De La Cruz, K. M., Noa-Copaja, S. J., Turpo-Gebera, O., Montesinos-Valencia, C. C., Bazán-Velasquez, S. M., & Pérez-Postigo, G. S. (2023). Use of gamification in English learning in higher education: A systematic review. *Journal of Technology and Science Education*, *13*(2), 480. https://doi.org/10.3926/jotse.1740

Li, X., Xia, Q., Chu, S. K. W., & Yang, Y. (2022). Using gamification to facilitate students self-regulation in E-Learning: A case study on students L2 English learning. *Sustainability*, *14*(12), 7008. https://doi.org/10.3390/su14127008

Lintangsari, A. P., & Emaliana, I. (2020). Inclusive education services for the blind: Values, roles, and challenges of university EFL teachers. *International Journal of Evaluation and Research in Education (IJERE)*, *9*(2), 439. https://doi.org/10.11591/ijere.v9i2.20436

Luo, Z. (2023). The effectiveness of gamified tools for Foreign Language Learning (FLL): A systematic review. *Behavioral Sciences*, *13*(4), 331. https://doi.org/10.3390/bs13040331

Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between level up and game over: A systematic literature review of gamification in education. *Sustainability*, *13*(4), 2247. https://doi.org/10.3390/su13042247

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research*, *283*, 112452. https://doi.org/10.1016/j.psychres.2019.06.027

Mulyadi, D., Wijayatiningsih, T. D., Singh, C. K. S., & Prastikawati, E. F. (2021). Effects of technology enhanced task-based language teaching on learners listening comprehension and speaking performance. *International Journal of Instruction*, *14*(3), 717–736. https://doi.org/10.29333/iji.2021.14342a

Nair, V., & Yunus, M. M. (2021). A systematic review of digital storytelling in improving speaking skills. *Sustainability*, *13*(17), 9829. https://doi.org/10.3390/su13179829

Namaziandost, E. (2019). The assessment of oral proficiency through holistic and analytic techniques of scoring: A comparative study. *Applied Linguistics Research Journal*, *3*(2), 70–82. https://doi.org/10.14744/alrj.2019.83792

Namaziandost, E., Shatalebi, V., & Nasri, M. (2019). The impact of cooperative learning on developing speaking ability and motivation toward learning English. *Journal of Language and Education*, *5*(3), 83–101. https://doi.org/10.17323/jle.2019.9809

Navidinia, H., Mobaraki, M., & Malekzadeh, F. (2019). Investigating the effect of noticing on EFL students speaking accuracy. *International Journal of Instruction*, *12*(1), 83–98. https://doi.org/10.29333/iji.2019.1216a

Orak, S. D. (2025). Turkish EFL teachers perspectives on AI-generated feedback: Negotiating trust, control, and pedagogical adaptation in writing instruction. *Applied Linguistics: Research, Measurement and Practice*, *1*(1), 74–94. https://doi.org/10.65334/n40kdg46

Ounis, A. (2017). The assessment of speaking skills at the tertiary level. *International Journal of English Linguistics*, *7*(4), p95. https://doi.org/10.1038/s41598-024-66047-4

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Quaglia, B. W. (2015). Planning for student variability: Universal design for learning in the music theory classroom and curriculum. *Music Theory Online*, *21*(1), 1–21. https://doi.org/10.30535/mto.21.1.6

Ramalingam, S., Yunus, M. M., & Hashim, H. (2022). Blended learning strategies for sustainable English as a second language education: A systematic review. *Sustainability*, *14*(13), 8051. https://doi.org/10.3390/su14138051

Rao, K., & Meo, G. (2016). Using universal design for learning to design standards-based lessons. *Sage Open*, *6*(4). https://doi.org/10.1177/2158244016680688

Rusconi, L., & Squillaci, M. (2023). Effects of a Universal Design for Learning (UDL) training course on the development teachers competences: A systematic review. *Education Sciences*, *13*(5), 466. https://doi.org/10.3390/educsci13050466

Saed, H. A., Haider, A. S., Al-Salman, S., & Hussein, R. F. (2021). The use of YouTube in developing the speaking skills of Jordanian EFL university students. *Heliyon*, *7*(7), e07543. https://doi.org/10.1016/j.heliyon.2021.e07543

Santhanasamy, C., & Yunus, M. M. (2021). A systematic review of flipped learning approach in improving speaking skills. *European Journal of Educational Research*, *11*(1), 127–139. https://doi.org/10.12973/eu-jer.11.1.127

Shen, M.-y., & Chiu, T.-y. (2019). EFL learners' English speaking difficulties and strategy use. *Education and Linguistics Research*, *5*(2), 88. https://doi.org/10.5296/elr.v5i2.15333

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2021). Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, *36*(3), 517–554. https://doi.org/10.1080/09588221.2021.1933540

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, *14*, 1–14. https://doi.org/10.3389/fpsyg.2023.1210187

Tai, T.-Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, *37*(5-6), 1281–1310. https://doi.org/10.1080/09588221.2022.2075013

Tai, T.-Y., & Chen, H. H.-J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. *Computer Assisted Language Learning*, *37*(5-6), 1224–1251. https://doi.org/10.1080/09588221.2022.2070219

Tavakoli, P., Nakatsuhara, F., & Hunter, A.-M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, *104*(1), 169–191. https://doi.org/10.1111/modl.12620

Tsou, W. (2005). Improving speaking skills through instruction in oral classroom participation. *Foreign Language Annals*, *38*(1), 46–55. https://doi.org/10.1111/j.1944-9720.2005.tb02452.x

Tuan, L. T. (2012). Teaching and assessing speaking performance through analytic scoring approach. *Theory and Practice in Language Studies*, *2*(4), 673–679. https://doi.org/10.4304/tpls.2.4.673-679

Yamani, H. A. (2021). A conceptual framework for integrating gamification in E-learning systems based on instructional design model. *International Journal of Emerging Technologies in Learning*, *16*(04), 14. https://doi.org/10.3991/ijet.v16i04.15693

Yan, W., Lowell, V. L., & Yang, L. (2024). Developing English language learners' speaking skills through applying a situated learning approach in VR-enhanced learning experiences. *Virtual Reality*, *28*(4), 167. https://doi.org/10.1007/s10055-024-01061-5

Yan, X., Kim, H. R., & Kim, J. Y. (2020). Dimensionality of speech fluency: Examining the relationships among Complexity, Accuracy, and Fluency (CAF) features of speaking performances on the Aptis test. *Language Testing*, *38*(4), 485–510. https://doi.org/10.1177/0265532220951508

Yenkimaleki, M., & van Heuven, V. J. (2019). The relative contribution of computer assisted prosody training vs. instructor based prosody teaching in developing speaking skills by interpreter trainees: An experimental study. *Speech Communication*, *107*, 48–57. https://doi.org/10.1016/j.specom.2019.01.006

Yusuf, A. (2025). Digital selves in the Nigerian second language classroom: Modelling the structural relationship between AI-enhanced identity expression, growth mindset, language self-efficacy, and second language grit. *Applied Linguistics: Research, Measurement and Practice*, *1*(1), 1–23. https://doi.org/10.65334/v2hm5720

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1030790

Zhang, W., Zhang, D., & Zhang, L. J. (2021). Metacognitive instruction for sustainable learning: Learners perceptions of task difficulty and use of metacognitive strategies in completing integrated speaking tasks. *Sustainability*, *13*(11), 6275. https://doi.org/10.3390/su13116275

Zhang, Z., & Crawford, J. (2023). EFL learners' motivation in a gamified formative assessment: The case of Quizizz. *Education and Information Technologies*, *29*(5), 6217–6239. https://doi.org/10.1007/s10639-023-12034-7

Zhussupova, R., & Shadiev, R. (2023). Digital storytelling to facilitate academic public speaking skills: Case study in culturally diverse multilingual classroom. *Journal of Computers in Education*, *10*(3), 499–526. https://doi.org/10.1007/s40692-023-00259-x

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in Artificial Intelligence (AI)-assisted language learning. *Sustainability*, *15*(4), 2872. https://doi.org/10.3390/su15042872

Appendix

Table III
Analytic speaking performance rubric for EFL learners

Criteria	5 – Excellent	4 – Good	3 – Satisfactory	2 – Limited	1 – Very Limited
Fluency	Speech flows naturally with minimal hesitation; pauses do not disrupt communication.	Generally smooth speech with occasional hesitation; meaning remains clear.	Noticeable pauses and hesitations, but communication is maintained.	Frequent hesitation and breakdowns that interrupt communication.	Very slow, fragmented speech; unable to maintain continuous utterances.
Accuracy (Grammar)	Consistently accurate structures; minimal or no errors; appropriate use of complex forms.	Mostly accurate grammar; errors do not impede meaning.	Frequent grammatical errors, but meaning remains understandable.	Errors are common and sometimes interfere with meaning.	Persistent errors that often obscure meaning; limited control of basic grammar.
Vocabulary Range	Wide and precise vocabulary used flexibly and appropriately.	Adequate range with some variety; occasional repetition or inappropriate use.	Mostly basic vocabulary; some inappropriate or repetitive choices.	Narrow vocabulary; frequent repetition; difficulty expressing intended meaning.	Very limited vocabulary; breakdowns due to lack of lexical resources.
Pronunciation	Clear, natural pronunciation; intelligible with appropriate stress and intonation.	Generally clear with minor issues; largely intelligible.	Noticeable mispronunciations but still understandable.	Frequent mispronunciations; intelligibility sometimes affected.	Pronunciation issues significantly hinder understanding.