Razi University

Technology Assisted Language Education

ISSN: 2981-1465

Homepage: https://tale.razi.ac.ir

Teacher-driven and AI-generated Dynamic Assessment and EFL Learners' Speaking Fluency: Exploring Perceptions via an Ecological Systems Theory (EST)

Reza Bozorg Aliabadi[1], Saman Ebadi[2]

[1]Department of English Language and Literature, Faculty of Humanities, Razi University, Kermanshah, Iran, rezaaliabadi2000@gmail.com

[2]Corresponding Author, Department of English Language and Literature, Faculty of Humanities, Razi University, Kermanshah, Iran. s.ebadi@razi.ac.ir

Article Info

Article type:
Research Article

Article history:
Received June 16, 2026
Received in revised form August 10, 2026
Accepted September 02, 2026
Published online September 24, 2026

Keywords:
Dynamic Assessment, speaking fluency, Ecological Systems Theory (EST), AI-generated feedback, teacher mediation, EFL learners, Mixed-methods

Abstract

This mixed-methods study, drawing on Ecological Systems Theory (EST), compared the contributions of two Dynamic Assessment (DA) modalities to speaking fluency among Iranian EFL learners. Participants included 60 B2-level male and female learners randomly assigned to a teacher-initiated DA group (n=30) and an AI-generated DA group (n=30). The former received teacher-driven DA based on Aljaafreh and Lantolf's graduated mediation, while the latter received feedback via the ChatGPT interface using the same protocol. Before and after the treatment, both groups took pre- and post-test IELTS speaking tests. One-way ANCOVA results showed that the teacher-driven DA group outperformed the AI-generated DA group. The paired-samples t-test results showed improvements in speaking fluency for both groups. The thematic analysis of transcribed data from semi-structured interviews, framed by EST, revealed learners' positive perceptions of teacher mediation; more specifically, learners perceived teacher mediation as offering attuned microsystemic mediation, institutional mesosystemic alignment, integrative exosystemic processes, and a culturally resonant macrosystem. In contrast, learners in the AI-generated DA perceived DA treatment as decontextualized; they found it mesosystemically disconnected and exosystemically fragmented. Learners also viewed AI-generated DA as culturally dissonant. These ecological contrasts explain why human mediation was superior, showing that despite the multiple benefits of AI tools, such as scalability, they fail to provide the multidimensional, context-sensitive scaffolding needed to embed learning within learners' authentic ecologies.

Cite this article: Bozorg Aliabadi, R., & Ebadi, S. (2026). Teacher-driven and AI-generated Dynamic Assessment and EFL Learners' Speaking Fluency: Exploring Perceptions via an Ecological Systems Theory (EST). *Technology Assisted Language Education*, *4*(3), 1-26. doi: 10.22126/tale.2026.14492.1196

©The Author(s). Publisher: Razi University

DOI: http//doi.org/10.22126/tale.2026.14492.1196

Introduction

Given the essential role of speaking in communicative competence and meaningful interaction, L2 instruction focuses on fostering this skill (Bygate, 1987; Renandya & Nguyen, 2022; Richards, 2008). In Iranian EFL education, research has identified multiple barriers that undermine speaking development. The main impediments include speaking-related anxiety, fear of negative evaluation, and underdeveloped grammar and vocabulary (Afshar & Asakereh, 2016; Hojati & Afghari, 2013). Moreover, ineffective curricula and limited educational resources compound these affective and linguistic challenges (Afshar & Asakereh, 2016). Mitigating these multifaceted obstacles requires interventions that use dynamic assessment (DA) and emerging technologies, such as Artificial Intelligence (AI).

Drawing on Vygotsky's sociocultural theory, DA links pedagogy and assessment by facilitating mediated interaction; indeed, DA emphasizes learners' potential rather than static ability (Lantolf & Poehner, 2004; Poehner, 2008). Previous studies have revealed the positive effects of DA across linguistic domains such as writing (e.g., Ebadi & Bashir, 2020; Ebadi & Rahimi, 2019; Kushki et al., 2022), reading comprehension (Al-Dawoody Abdulaal et al., 2022; Mohammadi & Babaii, 2022), listening (Ghenaat et al., 2022; Pileh Roud & Hidri, 2021), and grammatical knowledge (Barzegar & Azarizad, 2014; Davin, 2011). Regarding speaking skills, research (e.g., Ghahderijani et al., 2021; Rezaee et al., 2020; Safdari & Fathi, 2020) shows that ZPD-sensitive scaffolding improves learners' speaking fluency, accuracy, and complexity. Furthermore, research findings on technology-enhanced assessment, including computerized or mobile-based DA formats, have implications for fluency development (Ghahderijani et al., 2021; Rezaee et al., 2020). Despite these insights, the literature review reveals critical gaps. More specifically, limited research has examined speaking fluency, and most studies have compared DA to non-DA conditions; they have not explored the differential effects of various DA modalities. Recent efforts have targeted the integration of AI into DA associated with several skills, including writing (Akbari et al., 2026; Bashiri & Ebadi, 2024), reading comprehension (Leontjev et al., 2025; Sun et al., 2026), vocabulary (Jeon, 2023), and pragmatics (Saeedi & Soltani, 2025). However, this gap is not simply empirical. Prevailing DA models remain theoretically ill-equipped to account for AI's non-human agency within the ZPD. Since DA mediation is grounded in intersubjectivity and human intentionality, AI-generated prompts raise unresolved questions about whether algorithmic feedback can genuinely co-regulate learner performance or merely simulate scaffolding. This sociomaterial tension questions whether the ZPD can accommodate distributed, non-human mediation rather than assuming technological equivalence.

The growing adoption of AI for assessment has created multiple opportunities for personalized feedback and adaptive learning (Topping et al., 2025; Zhang et al., 2025), as evidenced by frameworks such as the Artificial Intelligence Assessment Scale (Perkins et al., 2024). However, most AI assessment applications focus on static assessment or intelligent tutoring rather than the dialogic, interactive mediation DA requires (Kolade et al., 2024; Perkins

et al., 2024). Newly developed designs, including agentic retrieval-augmented generation (RAG) systems, can provide iterative, rubric-aligned feedback cycles (Juárez et al., 2025); however, the literature review indicates that the capacity of such systems to scaffold speaking fluency within DA paradigms remains understudied. Presumably, current AI implementations in language assessment largely operate as asocial, decontextualized feedback engines, overlooking the socially negotiated, affectively attuned, and institutionally embedded nature of mediation. This ecological blind spot—where algorithmic prompts fail to cohere with learners' cultural expectations, institutional goals, and interpersonal trust—constitutes the central problem this study seeks to address. Validating AI tools that deliver real-time, graduated prompts compatible with learners' evolving responsiveness is highly challenging. These challenges are even more pronounced in diagnosing and scaffolding temporal variables such as pause duration and speech rate. Furthermore, studies on learners' perceptions of DA are limited. Equally uncertain is whether learners grant AI-generated mediation the same credibility, emotional salience, and developmental value they attribute to teacher-provided mediation.

Ecological Systems Theory (EST) (Bronfenbrenner, 1979) helps explain development within interconnected environmental contexts by emphasizing the interconnections between learning and multiple influences (e.g., immediate interactions and cultural norms) (Bronfenbrenner, 1979; Van Lier, 1997). Much existing DA research, however, has treated mediation as a decontextualized dyadic exchange, detached from the layered settings, tools, and cultural expectations that shape assessment. EST addresses this limitation by situating DA within interacting micro-, meso-, exo-, and macrosystems, thus showing how teacher and AI mediation are embedded in learners' immediate interactions and broader institutional and normative contexts. Therefore, this study compared the effects of teacher-driven DA versus AI-generated DA on speaking fluency among Iranian EFL learners. To do so, it used EST to shed light on learners' perceptions of the role teacher-mediated DA plays shaping their speaking performance.

The application of EST to understand how learner perceive DA enables them to examine the contributions of teacher scaffolding and AI tools- microsystemic interactions -to shaping learners' views. Moreover, EST sheds light on other variables, including mesosystemic connections between classroom assessment and institutional expectations, as well as macrosystemic, culturally mediated views toward technology and authority (Bronfenbrenner, 1979). To the best of my knowledge, no study has examined learners' perceptions of teacher-driven versus AI-generated DA from this multi-layered lens. Therefore, this study examines the differential effects of teacher-provided and AI-generated DA on speaking fluency, drawing on EST to explore learners' perceptions of the efficacy of teacher-mediated DA in shaping their speaking performance. To fill the lacuna in the literature, the following research questions were conceived:

> To what extent do teacher-driven and AI-generated DA differentially affect Iranian EFL learners' speaking fluency?

How do Iranian EFL learners perceive the efficacy of teacher-driven DA towards contributing to their speaking fluency from an EST perspective?

How do Iranian EFL learners perceive the efficacy of AI-generated DA towards contributing to their speaking fluency from an EST perspective?

Literature review

Dynamic Assessment and L2 Speaking Fluency

As a multidimensional ability, speaking fluency entails adequate speed, continuity, and automaticity and plays an essential role in successful communication (de Jong et al., 2013; Lennon, 1990; Segalowitz, 2010; Skehan, 2009). Speaking fluency is often measured in temporal variables, including speech rate, mean length of runs (MLU), and pause frequency (de Jong et al., 2013; Tavakoli & Skehan, 2005). Most EFL learners find fluent oral production challenging because multiple barriers, especially cognitive, linguistic, and affective ones, undermine their efforts to master this skill. This underscores the need for targeted instructional interventions (Skehan, 2009).

Dynamic assessment is underpinned by Vygotsky's (1978) zone of proximal development and integrates teaching and assessment through mediated dialogue. Indeed, DA emphasizes enhancing emergent abilities rather than static competence (Lantolf & Poehner, 2004; Poehner, 2008). Research attests to the positive effects of DA on L2 speaking; more specifically, DA has been found to improve fluency by providing ZPD-sensitive scaffolding. Intervention studies also document improvements in oral fluency and vocabulary retention (Safdari & Fathi, 2020; YarAhmadi & Kargar Behbahani, 2025). Moreover, research has revealed the positive effects of computerized and mobile-based DA formats on fluency development (Ghahderijani et al., 2021; Rezaee et al., 2020).

Before reviewing empirical findings, a conceptual distinction is necessary. DA is not a monolith but a continuum from interventionist mediation, which follows standardized prompts, to interactionist mediation, which relies on emergent, co-constructed dialogue. This distinction matters when comparing human and AI mediators, as teacher-driven DA tends toward the interactionist pole, whereas AI-generated DA necessarily operates closer to the interventionist pole, even when it uses graduated prompts. Most studies on DA and speaking fluency have compared DA with non-DA conditions, and comparisons among DA modalities remain underexplored (Safdari & Fathi, 2020; YarAhmadi & Kargar Behbahani, 2025). Furthermore, few, if any, studies have compared the effects of teacher-provided versus AI-generated mediation on different aspects of fluency, such as pause behavior and speech rate. This gap is critical, given the growing adoption of AI in L2 assessment and the need to understand how different mediation sources affect fluency outcomes.

AI-Generated Dynamic Assessment in Language Education

As a multidimensional ability, speaking fluency entails adequate speed, continuity, and automaticity and plays an essential role in successful communication (de Jong et al., 2013; Lennon, 1990; Segalowitz, 2010; Skehan, 2009). Researchers often measure speaking fluency using temporal variables, including speech rate, mean length of utterance (MLU), and pause frequency (de Jong et al., 2013; Tavakoli & Skehan, 2005). Most EFL learners find fluent oral production challenging because multiple barriers, especially cognitive, linguistic, and affective ones, undermine their efforts to master this skill. This underscores the need for targeted instructional interventions (Skehan, 2009).

AI-generated DA encompasses automated mediation formats, from rule-based tutoring systems to large language model (LLM)-driven feedback agents (Hung et al., 2024; Kim, 2026). In speaking assessment, such systems typically analyze temporal features—pause duration, articulation rate, and mean run length (Lu & Rogers, 2026)—and generate prompts based on predefined thresholds or learned models. Unlike human mediation, however, AI-generated scaffolds are constrained by training data and programmed response paths, making them more standardized than contingent (Rad, 2026; Raine, 2025). This contrast allows the study to investigate whether algorithmic mediation can approximate the contingency of human scaffolding rather than assuming equivalence. Most studies conducted on DA and speaking fluency have compared DA with non-DA conditions, and comparisons among DA modalities remain underexplored (Safdari & Fathi, 2020; YarAhmadi & Kargar Behbahani, 2025). This gap is critical, given the growing adoption of AI in L2 assessment and the need to understand how different mediation sources affect fluency outcomes.

An Integrated Socio-Ecological Framework for DA

The present study treats EST not as an add-on to SCT but as its ecological specification, recognizing that human development unfolds through dynamic interactions between individuals and nested environmental systems (Bronfenbrenner & Morris, 2007). In this synthesis, the microsystem is the interactional space where ZPD-sensitive mediation occurs: teacher or AI prompts are semiotic tools learners internalize through assisted performance (Hill & Ducasse, 2022). Mesosystemic links connect immediate mediation to classroom routines and curricula, determining whether DA scaffolds are reinforced or contradicted outside the tutorial (Sorayyaee et al., 2022). The exosystem exerts indirect influence through professional development, assessment policies, and AI design decisions, shaping which mediation forms are available and valued (Shirvan et al., 2021). Finally, the macrosystem transmits culturally mediated beliefs about authority, technology, and error correction, conditioning how learners interpret the credibility and emotional weight of human versus algorithmic feedback (Long & Xu, 2026).

Thus, microsystemic mediation is not merely situated in a context; each EST layer enables or constrains the quality of ZPD-based interaction. This integrated lens explains why the same graduated prompt may function interactionally in one setting and interventionally in another, depending on ecological affordances and constraints (Ge et al., 2026). The quality of ecological

coupling between AI, teachers, peers, and the learning environment mediates AI's actual effect on learner engagement (Curle, 2026), while institutional policies and cultural beliefs about language ability shape how feedback is received (Long & Xu, 2026). Understanding DA through EST thus reveals that scaffolding effectiveness depends not only on the prompt itself but on the full ecosystem in which it is embedded.

Learner Perceptions of DA and Ecological Systems Theory

Most studies on perceptions of DA (e.g., Afshari et al., 2020; Mohammadi & Babaii, 2022; Mohammadi et al., 2020; Parsaee & Akbarpour, 2022) have focused on teachers, examining their assessment literacy and perceptions of DA used for various skills. Overall, the findings indicate positive views but lack a robust conceptual framework and ignore learner perspectives. Indeed, the lack of attention to learners' perceptions and self-reflection regarding DA is a noteworthy gap given that learners are the primary recipients of mediation (Karimi & Shafiee, 2014).

Ecological Systems Theory (EST; Bronfenbrenner, 1979) provides a holistic framework for understanding learners' perceptions and evolving experiences within nested environmental systems. This covers a wide range of contexts, including immediate interactions (microsystem) and broader cultural values (macrosystem) (Bronfenbrenner & Morris, 2007). As Bronfenbrenner (1979) postulated, the microsystem refers to the immediate environment; the mesosystem characterizes settings beyond the immediate context; the exosystem refers to the relationships and processes between the previous layers; and the macrosystem refers to the realization of a specific culture. Researchers have applied EST to L2 learner variables such as anxiety, willingness to communicate, and motivation (Kasbi & Elahi Shirvan, 2017; Khajavy et al., 2016; Peng, 2012), and most findings attest to its explanatory power in accounting for layered influences on language-learning experiences.

Method

Participants

The initial participants comprised 78 male (N = 41) and female (N = 37) EFL learners selected through purposive sampling. They were enrolled at the B2 proficiency level and placed according to the Common European Framework of Reference for Languages (CEFR). The study selected this level because it focused on conversational proficiency, and all participants needed to converse effectively with the chatbot. Put simply, this level of proficiency was required for learners to interact with the chatbot. Moreover, because the study used English as the medium of instruction, it selected a proficiency level considered to have relatively high potential for engaging with the chatbot. The study adhered to ethical considerations by obtaining participants' written consent before the study began. In addition, all participants were briefed on the research process and aims, and they were assured that they could withdraw at

any time. Moreover, they were told that all data would be kept confidential and used solely for this research study.

Instruments

Oxford Quick Placement Test

The Oxford Quick Placement Test (OQPT), developed by Oxford University Press, consists of 60 items designed to evaluate language proficiency. This assessment categorizes learners' performance across a spectrum of proficiency levels, ranging from beginner to highly advanced: scores of 1-17 indicate a Beginner level, 18-27 signify Elementary, 28-36 denote Lower-Intermediate, 37-47 represent Upper-Intermediate, 48-55 correspond to Advanced, and scores of 56-60 indicate Highly Advanced proficiency. The scoring system is aligned with the Common European Framework of Reference (CEFR). For this study, we included only participants who scored between 37 and 47. We evaluated the OQPT's internal consistency using Cronbach's Alpha, yielding a coefficient of .87, which is considered acceptable.

IELTS Speaking Skill Test and Fluency Evaluation

This study drew on the oral evaluation approach introduced by Ebadi and Ebadijalal (2020), and assessed EFL learners' fluency using the public version of the IELTS band descriptors. The participants took pre- and post-tests on three sections of the IELTS speaking test. More specifically, these sections had the following characteristics: section one covered personal questions that served as a warm-up activity; section two assessed participants' ability to describe a specific occasion for two minutes after a one-minute reflection period. This assessment evaluated extended discourse and provided a window into learners' fluency in spontaneous speech. Section three addressed more abstract societal questions, eliciting participants' reflections on their previous responses and thus serving as a discussion segment. The criteria against which the IELTS test assessed learners' speaking skill were as follows: fluency and coherence (the extent to which a speaker can speak fluently and connect ideas smoothly), vocabulary (the speaker's ability to use vocabulary accurately and appropriately), grammatical range and accuracy (the correct application of a diverse range of grammatical structures), and pronunciation (the speaker's ability to use English phonetic features including intonation, stress patterns, and liaison in speech). Each domain was scored on a scale of 1 to 9 in accordance with the IELTS band descriptors. The overall speaking proficiency score was calculated by averaging the scores across all areas, with a focus on fluency in this research.

To ensure inter-rater reliability for fluency scores, the researchers recorded participants' speaking performance. Other EFL instructors experienced in using and interpreting the IELTS band descriptors then evaluated these recordings. This approach enabled comparison of fluency ratings across multiple raters, increasing the consistency and objectivity of the evaluations. Indeed, eliciting ratings from multiple instructors made it possible to identify and reduce potential biases. As a result, the ratings accurately reflected the participants' true fluency levels. Pearson's correlation coefficient was used to calculate inter-rater reliability, with the results

revealing an acceptable degree of agreement among raters. Overall, this quantitative analysis yielded robust evidence confirming the scoring system's reliability. Inter-rater reliability, as measured by Pearson's correlation coefficient, was 0.89 ($p < 0.01$) and 0.92 ($p < 0.01$) for the pretest and posttest, respectively. The rated recordings were the oral pretest and posttest responses from both groups. Both groups completed the same oral posttest, so fluency scores were based exclusively on spoken output. Thus, the reliability estimate reflects the same spoken outcome that was used to assess fluency gains.

Semi-structured Interviews

Two sets of semi-structured interview questions were developed to examine how Iranian EFL learners perceived the efficacy of teacher-driven DA (TDDA) and AI-generated DA (AIDA) in contributing to their speaking fluency from an EST perspective. To develop the interview questions for each set, the researchers first conducted an extensive literature review on teachers' and/or learners' perceptions of DA, AI assessment, and EST to create an initial list of questions. We included one general question to set the scene. Then, there were eight questions, each with two parts, seeking responses related to the four facets of EST. A panel of two Ph.D. holders in TEFL reviewed both lists of questions to support content validity. The panel then reviewed the lists and made some modifications. Next, we piloted each list of questions with five learners to address vagueness or ambiguity and improve readability and clarity. The final lists of questions were prepared after minor changes based on the learners' comments during piloting. Learners could choose between English and their mother tongue (Persian) for the interview sessions. Therefore, we also prepared a Persian version of the interview questions. To this end, the interview questions were translated into Persian, and two MA holders in translation studies with about 15 years of experience checked the initial translation. The translation experts were then asked to back-translate the Persian version of the questions into English. The panel then compared the translators' English versions with the original versions. Next, they revised the initial Persian translation and prepared the final Persian version of the questions. The expert panel also advised revisions to improve the clarity and readability of the Persian translation.

Procedure

Initially, 78 Iranian male and female EFL learners at the upper-intermediate level were selected and given the Quick Oxford Placement Test. Based on the results, 60 learners who scored at the B2 proficiency level were recruited. The selected learners were placed into two experimental groups. Both groups received a speaking pretest, and their spoken products were evaluated using the IELTS band-score fluency criteria. Following that, the interventions unfolded as follows:

Teacher-Driven DA Group

The teacher-driven DA intervention adapted Aljaafreh and Lantolf's (1994) graduated mediation framework to oral production. Mediation sequences progressed from implicit to

explicit scaffolding during speaking tasks (e.g., picture narratives, opinion exchanges). The following table details the modified mediation steps with speaking-specific examples:

Table 1

Teacher-Driven DA Mediation Protocol for Speaking Fluency

No	Mediation Description	Example (Student Utterance: *"Yesterday I go to market with my brother..."*)
1	Strategic pause + expectant gaze	Teacher remains silent after error while maintaining eye contact
2	Contextual repetition with rising intonation	"Yesterday you... go to market?"
3	Error-locator question	"Is there something unusual about the verb here?"
4	Time-frame narrowing	"Let's focus on the time word you used"
5	Grammatical category hint	"This might be about verb conjugation"
6	Contrastive recast	"You said 'go'... but when we talk about past, we usually...?"
7	Explicit error identification	"Remember: irregular past tense doesn't add -ed"
8	Guided self-correction prompt	"Try changing the form considering 'yesterday'"
9	Phonological clue	"The correct form sounds like /wen(t)/"
10	Modeled correction	"The standard way is: 'I went...'"
11	Metalinguistic explanation	"We use past simple for completed actions at specific past times"
12	Generative extension	"Now use this pattern in three new sentences about last week"

More specifically, the teacher-driven DA proceeded as follows: seeing a learner's struggle with speaking, e.g., producing a grammatically incorrect utterance, the teacher provided support by initiating a 12-step mediation sequence. This involved providing the most implicit feedback at the outset, then moving toward the most explicit level. For instance, if a learner made an erroneous sentence like "last week my friend come over," the teacher paused for a second and gazed (Step 1). Seeing no self-correction, the teacher proceeded to pose questions with rising intonation ("Last week, your friend... came over?" – Step 2); the teacher then posed error-locator questions (Step 3), revealing progressively more explicit cues like time-frame narrowing, accentuating grammatical hints, and contrastive recasts. If these initiatives led to no self-repair, the teacher resorted to modeled correction (Step 10) and metalinguistic explanation (Step 11). Ultimately, the teacher set up a generative extension task to guide the learner in producing new sentences using the corrected form (Step 12). Indeed, the teacher modified the mediation hierarchy based on the learner's responsiveness, ensuring a supportive, dialogic interaction aimed at fostering self-regulation in speaking fluency.

AI-Generated DA Group (AIDA)

The AI-driven DA group used the ChatGPT interface, which was carefully prompted to implement the same graduated mediation protocol used in the teacher-initiated DA condition.

This interface used a Generative Pre-trained Transformer (GPT) large language model and operated as a text-based conversational agent. The model version used was ChatGPT-4, accessed through OpenAI's text-based chat interface. To use the interface for this study, a detailed system prompt implemented AI-supported DA. This system produced the 12-step mediation sequence, with illustrative examples of learner utterances and appropriate responses at each level. For example, the prompt instructed the AI to start with the most implicit intervention—a silence prompt using ellipses and a questioning emoji to serve as a strategic pause. These questions and pauses intensified if the learner's next turn proved erroneous. For example, a learner's response like "Yesterday I go to market," triggered the first response from ChatGPT as follows: "Yesterday you... go to market?" 🤔 (mimicking contextual repetition). If there were no self-correction, the AI would proceed with posing an error-locator question like "Isn't this verb a bit weird here?" This was followed by time-frame narrowing ("Let's check out the time word you used"), grammatical category hints, and contrastive recasts. The AI resorted to the most explicit correction, such as "The standard form is 'I went...'", with a concise metalinguistic note. Then, the learner was encouraged to consolidate and internalize the new rule with a generative prompt such as "Now make three new sentences about last week using this pattern." The full system prompt, escalation script, and a sample AI–learner transcript illustrating Steps 1 to 12 are provided in Appendix.

To ensure adherence to the protocol, the AI was instructed to log each mediation step taken; that is, it could not skip implicit mediation to display explicit feedback unless the learner had already failed to self-correct after implicit prompts. Given ChatGPT's ability to process learner input in real time, the mediation unfolded immediately and consistently followed the escalation script. Moreover, the system retained flexibility within each step to accommodate varied learner productions. The learners engaged in real-time, individual interaction. They used the AI on laptops in a quiet laboratory setting and completed the same picture-narrative and opinion-exchange tasks as the TDDA group. Using a graduated approach, the AIDA supported the full spectrum of implicit-to-explicit mediation, with the AI adapting dynamically to accommodate linguistic accuracy and the learner's repair attempts, thereby adhering to the contingency principle central to DA. After the intervention, both groups took the same post-test and participated in the semi-structured interview procedures. This enabled the collection of fluency scores and rich qualitative data on learners' perceptions within the micro-, meso-, exo-, and macrosystems of EST.

Findings

Quantitative Results

To address the first research question, I ran a one-way ANCOVA. ANCOVA has several assumptions, including normality, covariate reliability, multicollinearity, linearity, homogeneity of regression, and homogeneity of variance, which must be met before running this statistical

test (Pallant, 2011). For the first assumption, Skewness and Kurtosis ratio values were examined. The respective results are presented in Table 2.

Table 2

Descriptive Statistics and Skewness and Kurtosis Values for Pretest and Posttest Scores

	N	Min	Max	Mean	SD	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pre Fluency TD	30	5.00	6.50	5.45	.37	.68	.42	.65	.83
Pre Fluency AI	30	5.00	6.00	5.68	.35	-.69	.42	-.69	.83
Post Fluency TD	30	5.50	7.00	6.46	.45	-.15	.42	-1.17	.83
Post Fluency AI	30	5.50	7.00	6.15	.32	.41	.42	.52	.83

As shown in Table 2, the Skewness and Kurtosis ratios (Statistic/Srd. Error) for the pretest and posttest scores fell within the range of ±1.96, indicating that the scores were normally distributed (Tabachnick & Fidell, 2007).

The second assumption, i.e., reliability of covariates, was assured via selecting a well-constructed and reliable instrument (Pallant, 2011) for measuring speaking fluency. The multicollinearity assumption was already met because there was only one covariate. As for the linearity assumption, the scatterplot of the variables was checked (Figure 1).

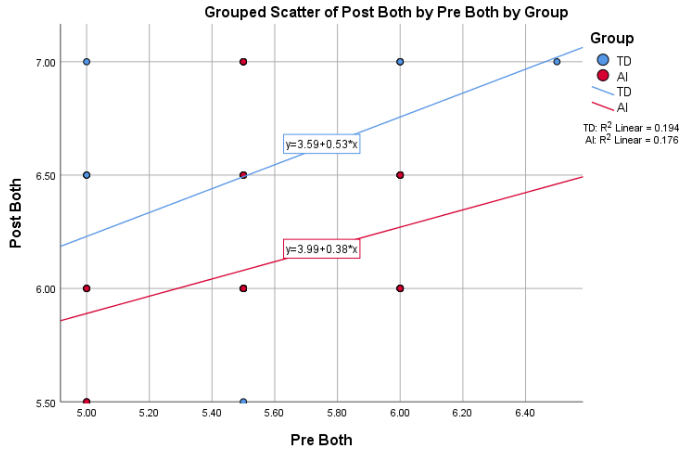


Figure 1. Scatterplot of pretest and posttest scores

As shown in Figure 1, the relationship between the dependent variable (Speaking Fluency posttest) and the covariate (pretest) formed a straight diagonal line, indicating a linear relationship. Thus, the assumption of linearity was not violated. To check homogeneity of regression slopes, the Tests of Between-Subjects Effects table was consulted. The respective results are presented in Table 3.

Table 3

Tests of Between-Subjects Effects for Pretest and Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.

Corrected Model	3.206a	3	1.069	8.154	.000
Intercept	3.634	1	3.634	27.724	.000
Group	.010	1	.010	.074	.787
Preboth	1.626	1	1.626	12.407	.001
Group * Preboth	.042	1	.042	.321	.573
Error	7.340	56	.131		
Total	2398.250	60			
Corrected Total	10.546	59			

As presented in Table 3, the significant value corresponding to Groups * Preboth equals .57, which is greater than 0.05, indicating that the assumption of the homogeneity of regression slopes was met. The last assumption was the homogeneity of variances, which was checked using Levene's test of variances (Table 4).

Table 4

Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
3.12	1	58	.082

Based on the results of Levene's test, variances in the dependent and covariate variables were equal; hence, the assumption of homogeneity of variances was met ($F = 3.12$, $p > .05$). After making sure that all of the assumptions were successfully met, the main ANCOVA output was examined. Table 5 displays the results of the ANCOVA.

Table 5

Results of ANCOVA for the Pretest and Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3.164a	2	1.582	12.216	.000	.300
Intercept	3.595	1	3.595	27.762	.000	.328
Preboth	1.660	1	1.660	12.817	.001	.184
Group	2.439	1	2.439	18.832	.000	.248
Error	7.382	57	.130			
Total	2398.250	60				
Corrected Total	10.546	59				

As shown in Table 5, the sig value for the groups was smaller than the critical value ($p < .001$), indicating a significant difference in speaking fluency scores between the two groups. The partial eta squared was .24, indicating a large effect size (Cohen, 1988). Table 6 displays the estimated marginal means for the speaking fluency scores.

Table 6

Estimated Marginal Means

			95% Confidence Interval	
Group	Mean	Std. Error	Lower Bound	Upper Bound
TD	6.520a	.067	6.385	6.655

AI	6.097a	.067	5.962	6.231

After controlling for pretest scores, the TD group had a higher adjusted speaking fluency mean (6.52) than the AI group (6.10). Next, paired-samples t-tests were run to examine the individual effects of TD dynamic assessment and AI dynamic assessment on speaking fluency. Table 7 displays the results.

Table 7

Results of Paired Samples T-test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	SD	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Pre Fluency TD - Post Fluency TD	-1.016	.44	.08	-1.18	-.85	-12.51	29	.00
Pair 2	Pre Fluency AI - Post Fluency AI	-.466	.36	.06	-.60	-.32	-6.91	29	.00

As shown in Table 7, both TD and AI dynamic assessments significantly improved speaking fluency from pretest to posttest ($p < .001$), with the TD group showing a larger mean increase (1.02) than the AI group (0.47).

Qualitative Results

The analysis used Braun and Clarke's (2006) thematic analysis. Although the EST provided a sensitizing framework, the researchers did not impose themes a priori. Initial coding was data-driven and resulted in descriptive codes such as "teacher waits for word search," "AI gives generic pause advice," and "exam relevance is explicit." Only after clustering did the researchers consider how these themes related to Bronfenbrenner's layers. Where data did not fit the ecological template, the authors report tensions within the relevant themes. This approach was intended to avoid the appearance of a perfectly clean theoretical mapping.

Learners' Perceptions of Teacher-Driven DA from an EST Perspective

Based on the thematic analysis procedures introduced by Braun and Clarke (2006), learners' perceptions of teacher-driven DA are shaped by four ecological themes as follows: (1) Attuned microsystemic mediation, (2) Institutional mesosystemic alignment, (3) Integrative exosystemic processes, and (4) Culturally resonant macrosystem.

Attuned microsystemic mediation

Learners found classroom interactions helpful, valuing the tailored scaffolding of fluency breakdowns that catered to their evolving knowledge. Moreover, the teacher considered learners' ZPD when providing scaffolding. In particular, the teacher's real-time diagnosis of pauses, which involved making a distinction between pauses caused by cognitive search and anxiety-induced silence, facilitated the creation of a psychologically safe microsystem. Ahmad

said: *"Whenever I came up with a tricky word, the teacher would first wait, then provide a first-sound hint. She worked out the cause of my silence. The teacher's supportive acceptance of my pause helped me speak more fluidly next time."* Other participants echoed this perception, including Zahra, who noted, "*The teacher knew exactly when my silence meant I was nervous versus when I was searching for vocabulary; that precision made me trust her more."*However, not all learners experienced this attunement uniformly. Sara, for example, described a mismatch between her needs and the teacher's wait-time: *"Sometimes she waited too long when I actually needed a word, and it made me more stuck. I wanted her to just give me the word."* This shows that microsystemic mediation, while often contingent, could also be perceived as misaligned.

Institutional mesosystemic alignment

In addition to the immediate interaction, teacher mediation was also perceived by learners as a bridge connecting speaking tasks with institutional expectations; that is, the mediations were consistent with curriculum aims and high-stakes exam criteria. Therefore, this explicit mesosystemic compatibility enhanced motivation. According to Sara: "*The teacher insisted on the connection between feedback and the speaking rubric used for exams: 'This pause and breakdown will make a dent in your fluency score, so go for this filler.' I could see for myself that the class activities and the teacher's mediations were in keeping with my institute's grading system*." This alignment was recurrently mentioned across the dataset. For instance, Hamid stated, "*Every time she corrected my fluency, she tied it to a specific exam criterion, which made me take the feedback seriously."*A minority of learners, however, found the constant exam linkage constraining. Leila noted, "*I know exams matter, but when she kept mentioning the rubric, I felt the fluency task became a test drill, not real speaking."* Thus, mesosystemic alignment was not uniformly motivating; for some, it reduced authenticity.

Integrative exosystemic processes

From an exosystemic perspective, learners acknowledged the teacher's efforts to coordinate microsystemic DA activities with mesosystemic institutional expectations. More specifically, the teacher sought to connect classroom mediation with both syllabus sequencing and exam preparation. This was considered of high importance by learners. Koroush noted: "*The teacher elaborated on the relevance of today's fluency activities to the unit topic and the final oral exam. Indeed, she helped us see the connections between the two worlds; otherwise, I'd feel like the sessions were far removed from my overall programme."* Several additional participants corroborated this integrative process. For example, Niloufar explained, *"She would remind us how each fluency task related to our portfolio and even to future course units, which made the feedback feel part of a bigger plan."* Amir added, "*The teacher linked our speaking practice to the curriculum map, so I understood why each session mattered for my overall progress."*Yet some learners did not perceive this integration. Babak commented, "The teacher talked about the syllabus, but I couldn't see the link to my weekly classes; maybe because I missed some

sessions." This suggests that exosystemic coordination was visible only to learners with regular attendance and could be missed by others.

Culturally resonant macrosystem

In the context of Iran's macrosystem, a teacher is considered an epistemic and moral authority. Thus, learners implicitly took the teacher's mediation for granted; they saw it as a culturally mandated role of mentorship. Mehrdad believed: *"The tips given by the teacher carry weight here. I embraced her prompts immediately as I was sure she knew my potential. My confidence in her support and authority helped me internalise corrections faster, which improved my fluency."*Other voices reinforced this cultural resonance. As Golnaz remarked, *"In our culture, we naturally respect the teacher's guidance; I accepted her feedback without questioning because that is how learning works here."*However, students did not always accept cultural authority uncritically. Neda stated, *"Because she is the teacher, I obeyed, but sometimes I felt her prompts were old-fashioned; I just didn't argue."* Cultural resonance, then, coexisted with quiet resistance, and some learners complied without genuine internalization.

Learners' Perceptions of AI-Generated DA from an EST Perspective

Based on the thematic analysis conducted on the learners' perceptions of AI-generated DA, the following four ecological themes emerged: (1) Decontextualized microsystemic interaction, (2) Mesosystemic disconnection, (3) Exosystemic fragmentation, and (4) Macrosystemic cultural dissonance.

Decontextualized microsystemic interaction

The analysis revealed that AI-generated feedback was immediate within the immediate microsystem; however, it was affectively and cognitively decontextualized. More specifically, the generated feedback could count pauses but failed to work out their cause. Put simply, the AI generated generic prompts. According to Delsa: *"The app asked me to 'reduce my pauses,' while I was searching in my mind for a word, not hesitating aimlessly. In contrast, the teacher can tell apart the causes of silence. The AI's judgment made me feel rushed, which made me more anxious instead of feeling supported."* This decontextualization was a recurring pattern across multiple participants. For instance, Behnam complained, *"The AI just told me to speak faster without realizing I was stuck on grammar, not speed."*Not all learners, however, viewed AI mediation as decontextualized. Pouya offered a contrasting account: *"The AI was quick, and I liked that it didn't judge me like a teacher; I could make mistakes without feeling embarrassed."* This positive case complicates the overall theme and suggests that perceived immediacy and anonymity could benefit some learners.

Mesosystemic disconnection

AI-mediated tasks were seen by learners as disconnected from their institutional mesosystem and of minimal relevance to syllabi and exam requirements. This perceived disconnection undermined engagement. Shahan said: "*It seemed that the AI topics were random. I mean I saw no relevance of the topics to my coursebook or the speaking test criteria. I wondered how these sessions could meet my institute's expectations, so I saw them as optional drills."* This sentiment was widely shared. As an instance, Nazanin observed, *"The AI tasks had no connection to what we were studying in class or to the exam format, so I didn't take them seriously."* Kaveh explained, *"I couldn't see how the AI feedback related to my course grade or the speaking rubric, which made it feel like a side activity."*Yet a minority found value despite disconnection. Raha noted, *"Even if it wasn't tied to my exam, it was extra speaking practice, and I felt my fluency improved in other contexts."* This suggests that mesosystemic disconnection did not completely negate perceived usefulness for all learners.

Exosystemic fragmentation

The AI-related mechanisms did not address the connection between AI microsystem interactions and broader institutional structures. More specifically, no mechanism integrated AI-generated feedback with progress tracking or assessment reporting. This led to a fragmented experience. Haniye observed: "*The teacher helped us see the connections between class activities and our exam criteria, while the AI just assigned marks. Indeed, I saw no relevance of these app sessions to my course grade or learning portfolio. I mean, it felt like a detached exercise."*In a similar vein, Babak noted, "The AI gave me a score but didn't explain how that score related to my overall performance or my institute's records."However, one learner noted that ChatGPT automatically saved logs, which she viewed as a form of progress tracking: "*I could scroll back and see my errors, so it wasn't fully detached for me"* (Sima). Therefore, fragmentation was perceived unevenly; for some, the artifact of saved interaction provided a minimal sense of continuity.

Macrosystemic cultural dissonance

From a macrosystemic perspective, learning in Iran is deeply interpersonal; however, AI lacked perceived legitimacy as an assessor in such a context. Influenced by cultural values, learners were more inclined to embrace human mentorship than machine monitoring. As a result, they remained skeptical of machine mediation. For example, Mahdi maintained: *"We're used to learning from a human master who empathizes with our struggle. It was hard for me to trust a machine given my shyness about speaking. Therefore, I often did not pay attention to its prompts, so it couldn't really help my fluency."* Several participants echoed this cultural skepticism. As Leila stated, *"In our culture, assessment comes from someone who knows you personally; a machine's feedback felt cold and irrelevant to my actual struggles."*Some learners, however, reported cultural acceptance of technology. Kaveh stated, *"I use AI for many things; I didn't feel it was against our culture, just different from a teacher."* Thus,

macrosystemic dissonance was not uniform across the sample, and cultural values interacted with individual dispositions toward technology.

Discussion

This study explored the possible differential contributions of teacher-provided and AI-generated dynamic assessment (DA) to speaking fluency among Iranian EFL learners. The study also examined how learners perceived each modality's efficacy using Ecological Systems Theory (EST). Quantitative analysis showed significantly stronger positive effects of teacher-provided DA on fluency than AI-generated DA, though both conditions improved fluency. The qualitative analysis drew on Bronfenbrenner's (1979) nested systems, which highlighted the role of ecological mechanisms in explaining this differential outcome. However, because the interview guide was developed from EST categories, the qualitative themes should be interpreted as an application of the theory to learner accounts rather than as independent evidence that the theory exists. The value of the analysis, therefore, lies not in confirming that the mesosystem or macrosystem matters, but in showing how these layers interact to shape the ZPD—specifically, how institutional and cultural conditions turn microsystemic mediation on or off.

One reason for the higher effects of teacher-driven DA is its ecological resonance across all four layers. From a microsystemic perspective, learners perceived the teacher's mediation as tailored to their ZPD. The teacher could distinguish between pauses stemming from cognitive word-searching and anxiety-induced hesitation. Accordingly, the teacher adjusted prompts based on the type of silence recognized. This dialogic adjustment is central to DA (Lantolf & Poehner, 2004; Poehner, 2008) and directly addresses the temporal and affective aspects of speaking fluency (de Jong et al., 2013; Skehan, 2009). Previous studies have demonstrated the positive effects of such tailored scaffolding on oral fluency (Safdari & Fathi, 2020; YarAhmadi & Kargar Behbahani, 2025). In the same vein, the qualitative data on participants' perceptions indicated that they found this attunement psychologically safe, alleviating the debilitating anxiety typical of Iranian EFL contexts (Afshar & Asakereh, 2016; Hojati & Afghari, 2013).

Drawing on Social Presence Theory, the teacher's physical co-presence, voice, and gaze may have lowered the affective filter in ways that text-based mediation cannot easily replicate. This social presence, rather than merely the mediation steps themselves, likely strengthened learners' willingness to respond to prompts. Beyond the immediate interaction, the results showed that teacher mediation aligned with institutional exam criteria from a mesosystemic perspective. This fostered learners' motivation, supporting previous research findings that alignment between classroom assessment and curricular goals increases perceived efficacy (Afshari et al., 2020; Mohammadi & Babaii, 2022). Moreover, this study offers a new insight from an exosystemic perspective: learners saw the teacher's efforts to integrate microsystemic DA sessions with mesosystemic structures (e.g., syllabus arrangements, high-stakes exam formats, and portfolio records). This act, known in the literature as integrative exosystemic processes, made the learners feel that the assessment was coherent and purposeful. Previous studies on DA perception have not addressed this dimension. However, the contribution is not

discovering that the exosystem exists; it is recognizing that the ZPD in oral fluency is ecologically distributed. When the teacher explicitly linked a moment-by-moment mediation to institutional outcomes, learners could internalize the prompt as meaningful rather than incidental, thereby expanding the ZPD beyond the immediate dyad. Regarding the macrosystemic layer, the culturally colored authority of the Iranian teacher (Afshar & Asakereh, 2016) fostered deep trust and a readiness to internalize corrective prompts. This trust-based yielding is deemed as a mechanism foundational to DA's developmental potential (Poehner, 2008). However, this is not a universal claim of human superiority; it is a claim about ecological fit in this specific context. The teacher's advantage depended on conditions—oral modality, institutional alignment, and cultural legitimacy—that may not transfer to other settings.

AI-generated DA contrasts with teacher-provided DA, and both conditions led to measurable improvement; however, learners perceived the former as ecologically incongruent across all four systems, which limited its efficacy. From a microsystemic perspective, learners perceived AI-generated feedback as decontextualized: although it could quantify pause frequency and speech rate, it failed to identify the invisible cognitive or affective causes of dysfluency. This finding lends credence to concerns that current AI tools cannot provide the real-time, contingent mediation needed to improve oral skills (Kolade et al., 2024; Perkins et al., 2024). The machine nature of AI fundamentally disrupts the intersubjective co-regulation central to Vygotskian mediation. Unlike a human mediator who shares intentionality and can enter the learner's zone of proximal development as a culturally attuned participant, ChatGPT operates through algorithmic pattern-matching and lacks genuine reciprocity, affective resonance, and negotiated meaning. This ontological discontinuity severs the bridge between microsystemic prompts and macrosystemic values, impeding scaffold internalization. Thus, while the teacher engaged in scaffolding—a relational, co-constructed process of guided participation—the AI merely issued prompts: algorithmically triggered, decontextualized cues lacking the dialogic reciprocity necessary for developmental transformation. Moreover, this discovery speaks to why AI-DA applications are mostly earmarked for written or receptive domains (Akbari et al., 2026; Bashiri & Ebadi, 2024; Jeon, 2023; Leontjev et al., 2025; Sun et al., 2026). At the mesosystemic level, the learners perceived the tasks as far removed from the syllabi and examination criteria, which diminished their perceived relevance.

This study also offers an exosystemic insight: the fragmentation theme signaled the exclusion of links between AI-generated feedback and learners' overall academic progress or institutional reporting; indeed, the AI was found to operate in isolation, avoiding the bridging of the micro and meso layers that the teacher did. This exosystemic fragmentation may have reduced transfer, but it does not fully explain the AI group's gain. Despite these ecological flaws, the AI group improved significantly. This pattern suggests a form of threshold efficacy: even decontextualized and institutionally disconnected mediation can provide enough structured input, repetition, and immediate feedback to support some fluency development. The AI may have functioned as a low-stakes practice environment in which learners could produce speech without face-threatening evaluation. From this perspective, AI-generated DA should not

be dismissed as inferior in all respects; rather, it may serve as a scalable supplement to, not a replacement for, human mediation. This interpretation aligns with emerging human-AI hybrid frameworks in language assessment.

From a macrosystemic perspective, the results indicated substantial cultural dissonance: in a learning culture that prioritizes human mentorship, AI was perceived to lack epistemic and moral legitimacy. Accordingly, learners ditched its prompts, mirroring conclusions from previous studies on trust-related challenges in AI-generated dynamic assessment (Akbari et al., 2026). This macrosystemic barrier likely deterred scaffold internalization, thereby compromising DA's core mechanism (Lantolf & Poehner, 2004). However, this macrosystemic interpretation should not be overstated. The sample comprised adult B2 learners with varied digital profiles, and some participants reported positive attitudes toward AI. The perceived dissonance may therefore reflect the clunky or text-heavy automated interface as much as any deep-seated cultural rejection of technology. Future research should disentangle interface effects from cultural values. As a result, AI-generated DA's microsystemic shortcomings were exacerbated by gaps at every subsequent ecological layer, resulting in its inferior fluency outcomes. This study suggests that such gaps limited but did not eliminate AI-mediated gains.

Overall, these findings show that both feedback immediacy (i.e., adaptability) and ecological fit contribute to the differential effect of DA modalities on speaking fluency. Moreover, the in-depth analysis of learners' perceptions across all four EST layers revealed teacher-provided mediation as a superior contributor to enhancing complex oral production in this context, though AI's partial utility was also confirmed. The theoretical contribution is not that EST categories exist in the data, but that the ZPD in oral fluency is jointly shaped by microsystemic contingency, mesosystemic alignment, exosystemic coherence, and macrosystemic legitimacy. This reframes dynamic assessment from a purely dyadic interaction to an ecologically embedded process. These findings extend EST by showing that AI is not a novel ecological layer; rather, it operates as a non-integrated artifact within the microsystem that lacks the systemic permeability needed to traverse meso-, exo-, and macrosystemic boundaries. This challenges the implicit assumption that technological tools automatically enter the developmental ecology and calls for theorizing non-human agents as contingent affordances whose efficacy depends on their capacity to achieve ecological resonance. This study contributes to DA literature by directly comparing the contributions of two modalities to speaking fluency—a long-standing gap (Safdari & Fathi, 2020; YarAhmadi & Kargar Behbahani, 2025). The findings suggest a wider range of EST applications; that is, the theory can also be applied to the previously unaddressed domain of assessment perceptions. Future studies should avoid circularity by using inductive thematic analysis in which ecological categories emerge a posteriori, or by mixing EST-informed and open-ended prompts. They should also examine voice-based AI mediation under conditions of higher social presence and lower latency to determine whether the observed gains can be amplified.

Conclusion

This study showed the positive effects of both teacher-provided and AI-generated dynamic assessment (DA) on Iranian EFL learners' speaking fluency; however, teacher-provided DA was found to yield significantly greater gains. Drawing on an Ecological Systems Theory lens, the study analyzed qualitative results regarding learners' perceptions. The findings indicated that learners perceived teacher mediation as ecologically congruent across the four layers. Furthermore, AI-generated mediation was seen as decontextualized, disconnected, fragmented, and culturally dissonant. These perceptions reveal the mechanisms underlying the quantitative differential effect.

This study has important pedagogical implications; the findings substantiate the crucial role of teacher mediation in addressing the affective and cognitive challenges associated with oral fluency. Teachers need to consider dialogic, ZPD-sensitive scaffolding when mediating DA activities (Lantolf & Poehner, 2004). Meanwhile, they should selectively integrate AI as a supplementary tool for practice and preliminary feedback. Indeed, teachers should consider ecological fit so that curricula and institutional goals align with AI-generated mediation. Doing so would foster an AI-teacher synergy that combines immediacy with human empathy (Akbari et al., 2026; Bashiri & Ebadi, 2024).

As with other studies, this study is constrained by several limitations. The study's setting was a single Iranian language institute, which limits generalizability; the relatively small sample size may also have constrained statistical power. Moreover, the AI tool used a rule-based pause-detection algorithm rather than an advanced agentic RAG system (Juárez et al., 2025); therefore, the intervention duration may not have captured long-term development. Lastly, learners' perceptions were self-reported, which could have been influenced by cultural desirability bias.

Prospective studies could replicate the comparative design across different cultural contexts, as the current study used purposive sampling within a single language institute and a specific proficiency level. Accordingly, the results cannot be generalized to various cultural or institutional contexts and proficiency levels. Moreover, longitudinal, mixed-methods studies would be highly insightful for capturing online cognitive and affective responses to mediation, using process-tracing measures such as eye-tracking and stimulated recall. Advanced generative AI systems capable of recognizing speech emotion and offering adaptive dialogue could bridge the ecological gaps identified here; they could examine the effects of these systems and their capabilities within a genuine DA framework.

Bio-data

First Author: is a Ph.D. candidate in the Department of English Language and Literature, Razi University, Kermanshah. He is interested in conducting research in CALL, dynamic assessment, feedback, teacher education, writing, and affective variables in SLA. He has published in Current Psychology, Journal of Writing Research, International Journal of

Educational Research Open, International Journal of Language Testing, and Language Related Research. He has also given presentations at national conferences.

Second Author: a professor of applied linguistics at the Department of English Language and Literature, Razi University, Kermanshah. He is interested in conducting research in CALL, MALL, dynamic assessment, Artificial intelligence, and SLA. He has published in many distinguished national and international journals. He has also given presentations at national and international conferences.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest: The authors declare that they have no competing interests.

References

Afshar, H. S., & Asakereh, A. (2016). Speaking skills problems encountered by Iranian EFL freshmen and seniors from their own and their English instructors' perspectives. *Electronic Journal of Foreign Language Teaching*, *13*(1), 112–130.

Afshari, H., Amirian, Z., & Tavakoli, M. (2020). Applying group dynamic assessment procedures to support EFL writing development: Learner achievement, learners' and teachers' perceptions. *Journal of Writing Research,11*(3), 445-476. https://doi.org/10.17239/jowr-2020.11.03.02

Akbari, F., Amirian, S. M. R., Zareian, G., & Ghaniabadi, S. (2026). The Effect effect of Artificial artificial Intelligence intelligence and Dynamic dynamic aAssessment sSynergy on wWriting dDevelopment of Iranian nNursing sStudents. *International Journal of Language Testing*, *16*(1), 84-99. https://doi.org/10.22034/ijlt.2025.546154.1480

Al-Dawoody Abdulaal, M. A., Ramadan Khalil, N., Heji Alenazi, M., & Robso Wodajo, M. (2022). Dynamic vs non-dynamic assessments: Impacts on intermediate EFL learners' receptive skills. *Education Research International, 3,* 1-12. https://doi.org/10.1155/2022/5372929

Barzegar, R., & Azarizad, R. (2014). Using dynamic assessment to improve L2 learners' knowledge of grammar: Evidence from the tenses. In M. Pawlak & J. Bielak (Eds.), *Classroom-oriented research: Achievements and challenges* (pp. 219–227). Springer.

Bashiri, S., & Ebadi, S. (2024). The impact of AI-based group dynamic assessment on L2 writing proficiency. *Technology Assisted Language Education*, *2*(3), 151-173. https://doi.org/10.22126/tale.2025.11488.1081

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical*

models of human development (6th ed., pp. 793–828). Wiley. https://doi.org/10.1002/9780470147658.chpsy0114

Bygate, M. (1987). *Speaking*. Oxford University Press.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences (2nd ed)*. Lawrence Erlbaum Associates.

Curle, S. M. (2026). Beyond automated judgement: reframing Intelligent Computer-Assisted Language Assessment (ICALA) as an assessment ecology. *Language Testing in Asia*, *16*(1), 26. https://doi.org/10.1186/s40468-026-00436-6

Davin, K. J. (2011). *Group dynamic assessment in an early foreign language learning program: Tracking movement through the zone of proximal development* (Doctoral dissertation). University of Pittsburgh, Pittsburgh, Pennsylvania.

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34*(5), 893–916. https://doi.org/10.1017/S0142716412000069

Ebadi, S., & Bashir, S. (2020). An exploration into EFL learners' writing skills via mobile-based dynamic assessment. *Educational Information Technologies,* 26. https://doi.org/10.1007/s10639-020-10348-4

Ebadi, S., & Ebadijalal, M. (2020). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning, 25*, 1-25. https://doi.org/10.1080/09588221.2020.1854311

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning, 32*(5–6), 527–555. https://doi.org/10.1080/09588221.2018.1527362.

Ebadi, S., Aliabadi, R. B., & Ajabshir, Z. F. (2024). EFL teachers' engagement-promoting strategies in online gamification: An ecological systems theory (EST) perspective. *Current Psychology, 43*(28), 23713-23728. https://doi.org/10.1007/s12144-024-06143-w

Ge, L. L., Xia, Y., & Yang, C. (2026). Beyond tool use: How ecological coupling configures AI's empowerment for language learning engagement. *Frontiers in Psychology, 17*, 1747355. https://doi.org/10.3389/fpsyg.2026.1747355

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), 25. https://doi.org/10.1186/s40468-021-00144-3

Ghenaat, M., Esfahani, F. R., Shafiee, S., & Sepehri, M. (2022). Effect of interactionist, group and computerized dynamic assessment on Iranian EFL learners' listening comprehension: Impact of interactionist, group and computerized dynamic assessment. *Argentinian Journal of Applied Linguistics-ISSN 2314-3576*, *10*(1), 63-78.

Hill, K., & Ducasse, A. M. (2022). Contextual variables in written assessment feedback in a university-level Spanish program. *Studies in Language*, *11*(1), 17. https://doi.org/10.58379/VRFA3279

Hojati, A., & Afghari, A. (2013). An investigation of speaking-associated problems from students. *The Iranian EFL Journal*, *9* (1), 9-31. www.academia.edu/download/78926877/A_Cross-Disciplinary_Analysis_of_Themati20220116-18769-aq7sgp.pdf#page=9

Hung, J.-T., Cui, C., Popescu, D. M., Chatterjee, S., & Starner, T. (2024). Socratic Mind: Scalable oral assessment powered by AI. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale (L@S '24)* (pp. 1–6). ACM. https://doi.org/10.1145/3657604.3664661

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, *36*(7), 1338-1364. https://doi.org/10.1080/09588221.2021.1987272

Juárez, R., Hernández-Fernández, A., de Barros-Camargo, C., & Molero, D. (2025). Dynamic assessment with AI (Agentic RAG) and iterative feedback: A model for the digital transformation of higher education in the global EdTech ecosystem. *Algorithms*, *18*(11), 712. https://doi.org/10.3390/a18110712

Karimi, M., & Shafiee, Z. (2014). Iranian EFL teachers' perceptions of dynamic assessment: Exploring the role of education and length of service. *Australian Journal of Teacher Education, 39*(8). http://dx.doi.org/10.14221/ajte.2014v39n8.10

Kasbi, S., & Elahi Shirvan, M. (2017). Ecological understanding of foreign language speaking anxiety: Emerging patterns and dynamic systems. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(2), https://doi.org/10.1186/s40862-017-0026-y

Khajavy, G. H., Ghonsooly, B., Hosseini Fatemi, A., & Choi, C. W. (2016). Willingness to communicate in English: A microsystem model in the Iranian EFL classroom context. *TESOL Quarterly, 50*(1), 154-180. https://doi.org/10.1002/tesq.204

Kim, M. (2026). Custom GPT as mediator: Dynamic Assessment with beginner KFL learners. *Language Learning & Technology, 30*(1), 1–27. https://doi.org/10.64152/10125/73669

Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon*, *10*(4). https://www.cell.com/heliyon/fulltext/S2405-8440(24)01984-4

Kushki, A., Nassaji, H., Rahimi, M., (2022). Interventionist and interactionist dynamic assessment of argumentative writing in an EFL program, *System,* https://doi.org/10.1016/j.system.2022.102800.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics, 1*(1), 49–72. https://doi.org/10.1558/japl.1.1.49.55872

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Leontjev, D., Poehner, M., Huhta, A., & Pollari, P. (2025). Reading with a Chatbot-The added value of Generative AI as a resource in mediating learners in Dynamic Assessment of L2 English reading. *Studies in Language Assessment*, (2). https://doi.org/10.58379/LSFJ4641

Long, Z., & Xu, J. (2026). An ecological approach to L2 learners' engagement with written feedback. *Assessing Writing*, *68*, 101028. https://doi.org/10.1016/j.asw.2026.101028

Lu, J., & Rogers, J. (2026). Evaluating and enhancing the accuracy of automated fluency annotation tools in L2 research. *Research Methods in Applied Linguistics*, *5*(1), 100302. https://doi.org/10.1016/j.rmal.2026.100302

Mohammadi, S., & Babaii, E. (2022). The dynamic assessment of reading comprehension: An exploration of EFL teachers' perception and practice. *Journal of Modern Research in English Language* Studies*, 9*(2), 123-148. http://dx.10.30479/jmrels.2020.12998.1607

Pallant, J. S. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows (4th ed)*. Allen & Unwin, Crows Nest.

Parsaee, M., & Akbarpour, L. (2022). The relationship between Iranian EFL teachers' assessment literacy and their attitudes toward dynamic assessment. *Journal of Studies in Learning and Teaching English*, *11*(1), 1-20. https://sanad.iau.ir/Journal/jslte/Article/1106537

Peng, J. (2012). Towards an ecological understanding of willingness to communicate in EFL classrooms in China. *System, 40*, 203–213. https://doi.org/10.1016/j.system.2012.02.002

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, *21*(6), 49-66 https://doi.org/10.53761/q3azde36

Pileh Roud, L. F., & Hidri, S. (2021). Toward a sociocultural approach to computerized dynamic assessment of the TOEFL iBT listening comprehension test. *Education and Information Technologies*, *26*(4), 4943-4968. https://doi.org/10.1007/s10639-021-10498-z

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer.

Rad, H. S. (2026). Effects of dialogic and monologic AI/non-AI feedback on L2 speaking development: Exploring learner achievements, feedback literacy, and interactional moves. *ReCALL*, 1-18. https://doi.org/10.1017/S095834402610055X

Raine, P. (2025). Using GPT-3 as a More Knowledgeable Other to Provide Feedback on English Language Learners' Spoken Transcripts. *TESL-EJ*, *29*(2), 1-5. https://files.eric.ed.gov/fulltext/EJ1482976.pdf

Renandya, W. A., & Nguyen, M. T. T. (2022). Teaching speaking in L2 contexts. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 269–280). Routledge.

Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2020). Mobile-based dynamic assessment and the development of EFL students' oral fluency. *International Journal of Mobile Learning and Organization*, *14*(4), 511-532. https://doi.org/10.1504/IJMLO.2020.110789

Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge University Press.

Saeedi, Z., & Soltani, M. (2025). Developing an AI chatbot for language pragmatics instruction: From algorithms to dynamic assessment in an EFL context. *Computer Assisted Language Learning*, 1-22. https://doi.org/10.22126/tale.2025.11488.1081

Safdari, M., & Fathi, J. (2020). Investigating the role of dynamic assessment on speaking accuracy and fluency of pre-intermediate EFL learners. *Cogent Education*, *7*(1), 1818924. https://doi.org/10.1080/2331186X.2020.1818924

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Shirvan, M. E., Lou, N. M., & Taherian, T. (2021). Where do language mindsets come from? An ecological perspective on EFL students' mindsets about L2 writing. *Journal of Psycholinguistic Research*, *50*(5), 1065-1086. https://doi.org/10.1007/s10936-021-09787-y

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Sorayyaee, L., Shirvan, M. E., & Akbari, O. (2022). An ecological exploration of Iranian English as a foreign language learners' goal-directed visions. *Iranian Journal of Language Teaching Research*, *10*(1), 81-99. https://doi.org/10.30466/ijltr.2022.121123

Sun, J., Huang, Y., Chang, Y., Li, F., & Zhao, J. (2026). AI-assisted dynamic assessment design in lexical inference and reading comprehension. *Reading and Writing*, 1-23. https://doi.org/10.1007/s11145-026-10792-7

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics*. Pearson Education.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins.

Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education*, *22*(1), 3. https://doi.org/10.1186/s41239-024-00501-1

Van Lier, L. (1997). Observation from an ecological perspective. *TESOL Quarterly, 31*, 783–787.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

YarAhmadi, M., & Kargar Behbahani, H. (2025). The effect of interventionist dynamic assessment on Iranian EFL learners' vocabulary learning and retention: A sociocultural inquiry. *Language Testing in Asia*, *15*(1), 2. https://doi.org/10.1186/s40468-024-00337-6

Zhang, H., Perry, A., & Lee, I. (2025). Developing and validating the artificial intelligence literacy concept inventory: An instrument to assess artificial intelligence literacy among middle school students. *International Journal of Artificial Intelligence in Education*, *35*(1), 398-438. https://doi.org/10.1007/s40593-024-00398-x

Appendix

Step	**Mediation Description**	**ChatGPT Prompt**	**Example ChatGPT Response to Learner Utterance: "Yesterday I go to market with my brother..."**
1	Strategic pause + expectant gaze	Output only an ellipsis and a questioning emoji to signal a silent pause. Do not provide any correction.	"… 🤔" / "...
2	Contextual repetition with rising intonation	Repeat the learner's utterance up to the error, leaving the erroneous part with rising intonation. Use a question mark. Do not correct.	"Yesterday you... go to market?"
3	Error-locator question	Ask a question that directs attention to the location of the error without naming the error.	"Is there something unusual about the verb here?"
4	Time-frame narrowing	Direct the learner's attention to the time expression used in the utterance.	"Let's focus on the time word you used."
5	Grammatical category hint	Give a hint about the grammatical category involved in the error.	"This might be about verb conjugation."
6	Contrastive recast	Contrast the learner's form with the target form indirectly. Do not supply the full correction.	"You said 'go'... but when we talk about past, we usually...?"
7	Explicit error identification	Identify the error explicitly, but do not provide the corrected form.	"Remember: irregular past tense doesn't add -ed."
8	Guided self-correction prompt	Prompt the learner to self-correct using a contextual clue from the utterance.	"Try changing the form considering 'yesterday'."
9	Phonological clue	Provide a phonological hint about the correct spoken form.	"The correct form sounds like /wen(t)/."
10	Modeled correction	Provide the standard corrected form.	"The standard way is: 'I went...'"
11	Metalinguistic explanation	Provide a brief metalinguistic rule.	"We use past simple for completed actions at specific past times."
12	Generative extension	Ask the learner to produce three new sentences using the corrected pattern.	"Now use this pattern in three new sentences about last week."