

The Impact of Learning-Oriented Assessment in a Technology-Enhanced Environment on Iranian EFL Learners' Speaking Skills

Hamed. Fathi¹, Ali Asghar. Yousefi Azarfam^{2*}, Nesa. Nabifar³

¹ PhD Candidate, Department of English, TA.C., Islamic Azad University, Tabriz, Iran

² Assistant Professor in ELT, IAU Editorial assistant, JELPP

³ Assistant Professor in ELT, IAU

* Corresponding author email address: tabrizjal@gmail.com

Article Info

Article type:

Original Research

How to cite this article:

Fathi, H., Yousefi Azarfam, A. A., & Nabifar, N. (2026). The Impact of Learning-Oriented Assessment in a Technology-Enhanced Environment on Iranian EFL Learners' Speaking Skills. *Iranian Journal of Educational Sociology*, 9(2), 1-16.
<https://doi.org/10.61838/kman.ijes.1450>



© 2026 the authors. Published by Iranian Association for Sociology of Education, Tehran, Iran. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Purpose: This study aimed to examine whether implementing Learning-Oriented Assessment (LOA) in a technology-enhanced environment leads to significantly greater improvement and retention in Iranian EFL learners' speaking performance compared to non-technology LOA and traditional assessment conditions.

Methods and Materials: A mixed-methods, quasi-experimental design was employed with three groups: Technology-Enhanced LOA (TE-LOA), LOA-Only, and a traditional assessment control group. Sixty intermediate EFL learners were selected using the Oxford Quick Placement Test to ensure baseline comparability. Participants completed a speaking pretest, posttest, and delayed posttest over the course of the intervention. Speaking performance was evaluated using an analytic rubric measuring fluency, accuracy, pronunciation, and lexical resources, with acceptable inter-rater reliability established. Quantitative data were analyzed using mixed-design repeated-measures ANOVA to assess within-group changes over time and between-group differences, followed by Bonferroni-adjusted post-hoc comparisons.

Findings: The results revealed a statistically significant main effect of Time and a significant Group $\times$ Time interaction ($p < .001$), indicating differential development across conditions. The TE-LOA group demonstrated significantly greater gains in overall speaking performance compared to both the LOA-Only and control groups at posttest and delayed posttest ($p < .001$), with large effect sizes. Within-group comparisons showed significant improvements from pretest to posttest and sustained performance at delayed posttest in the TE-LOA group ($p < .001$), while the LOA-Only group showed moderate gains and the control group exhibited minimal change. Subskill analyses confirmed significant improvements in fluency, accuracy, and lexical complexity for the TE-LOA group ($p < .001$).

Conclusion: The findings indicate that integrating LOA with technology-enhanced environments significantly improves EFL learners' speaking performance and supports short-term retention of gains, suggesting that technology-mediated LOA provides a more effective and sustainable assessment framework than traditional or non-technology LOA approaches.

Keywords: Learning-Oriented Assessment, Technology-Enhanced Learning, Speaking Skills, EFL Learners, Formative Assessment, Language Assessment Literacy

1. Introduction

In recent decades, the conceptualization of assessment in second language education has undergone a substantial paradigm shift from a predominantly summative, score-oriented enterprise toward a more formative, learning-driven process embedded within instruction. Early classroom research emphasized the centrality of pedagogy and interaction in shaping learning outcomes, highlighting that assessment cannot be meaningfully separated from teaching practices and classroom ecology (Allwright & Bailey, 1991; Cortazzi & Jin, 1999). Building on these foundations, contemporary perspectives increasingly conceptualize assessment as an integral component of the learning process, one that actively mediates learners' cognitive engagement, self-regulation, and language development rather than merely documenting achievement (Carless, 2007, 2015a). This reconceptualization is particularly salient in English as a Foreign Language (EFL) contexts, where communicative competence and performance-based skills such as speaking require continuous practice, feedback, and iterative refinement rather than one-shot evaluation.

Within this evolving landscape, Learning-Oriented Assessment (LOA) has emerged as a theoretically robust and pedagogically consequential framework that seeks to align assessment practices with learning processes. LOA integrates formative and summative functions by emphasizing assessment tasks that promote learning, active learner engagement in evaluation processes, and feedback mechanisms that inform future performance (Carless, 2015b; Jones & Saville, 2016; Turner & Purpura, 2016). In this approach, assessment is not an endpoint but a cyclical process involving task performance, feedback reception, reflection, and re-performance. Such a perspective resonates with broader developments in language pedagogy, including sociocultural and interactionist theories that emphasize the role of feedback, noticing, and negotiated interaction in second language acquisition (Gass & Mackey, 2000, 2016). Consequently, LOA has been increasingly advocated as a means of enhancing both immediate performance and long-term learning capacity.

The effectiveness of LOA is closely tied to the development of language assessment literacy among teachers and learners. Assessment literacy encompasses the knowledge, skills, and principles required to design, implement, and interpret assessment in ways that support learning (Farhady & Tavassoli, 2021; Fulcher, 2021). In many EFL contexts, however, gaps in assessment literacy

have been identified as a significant barrier to the successful implementation of innovative assessment practices. Empirical studies have shown that teachers often experience tensions between theoretical principles and classroom realities, leading to discrepancies between intended and enacted assessment practices (Aliakbari et al., 2023; Rezagah, 2022). These challenges are further compounded by institutional constraints, high-stakes testing regimes, and limited professional development opportunities, which collectively reinforce traditional, exam-oriented approaches to assessment.

In the Iranian EFL context, these issues are particularly pronounced. Despite policy-level emphasis on communicative competence, classroom assessment practices often remain dominated by summative testing, with limited integration of formative, feedback-oriented processes (Muhammadpour & Khalili Sabet, 2025; Riazi & Razavipour, 2011). Research has consistently highlighted a gap between teachers' awareness of LOA principles and their actual classroom implementation, pointing to structural, contextual, and cognitive barriers (Derakhshan & Ghiasvand, 2022; Estaji & Kogani, 2024; Jalilzadeh, 2023). For instance, teachers may recognize the value of peer assessment and feedback but struggle to operationalize these practices within time-constrained and exam-driven environments. Similarly, learners may lack the metacognitive skills required to engage effectively in self-assessment, thereby limiting the potential benefits of LOA.

The emergence of digital technologies has introduced new possibilities for addressing these challenges and enhancing assessment practices. Technology-enhanced assessment environments enable the collection, storage, and analysis of performance data in ways that were previously impractical, thereby facilitating continuous monitoring and feedback (Burr et al., 2023; Jisc, 2010). In language learning, digital tools such as learning management systems, e-portfolios, and AI-driven applications can support iterative cycles of performance and feedback, making assessment more dynamic and learner-centered. For example, recent research has demonstrated that AI-powered conversational agents can enhance learners' speaking skills while simultaneously reducing anxiety, highlighting the potential of technology to create supportive and interactive learning environments (Ding & Yusof, 2025). Similarly, e-portfolio systems have been shown to promote reflective learning and sustained engagement with assessment criteria, thereby aligning closely with LOA principles (Rezadoust Siah Khaleh Sar et al., 2024).

Speaking, as a core component of communicative competence, presents unique challenges and opportunities for assessment. Unlike written skills, speaking is transient, interactive, and context-dependent, making it difficult to capture and evaluate using traditional assessment methods. Effective speaking assessment requires attention to multiple dimensions, including fluency, accuracy, pronunciation, and interactional competence, all of which develop through repeated practice and feedback (Byram, 1997; Gass & Mackey, 2016). LOA is particularly well-suited to this domain because it emphasizes ongoing assessment cycles and learner engagement, which are essential for developing complex, real-time language skills. Empirical evidence supports the effectiveness of LOA in improving speaking-related outcomes, including pronunciation and communicative competence (Ashegh Navaie, 2018; Estaji & Safari, 2023).

Technology can further enhance LOA in speaking by providing tools for recording, replaying, and analyzing oral performances, thereby making learning processes more visible and accessible. Social media platforms and digital communication tools have also been shown to facilitate language practice and feedback, contributing to improvements in pronunciation and fluency (Xodabande, 2017). Moreover, online and blended learning environments, which have become increasingly prevalent in the wake of the COVID-19 pandemic, have underscored the importance of flexible and adaptive assessment practices. Studies conducted during this period have highlighted both the challenges and opportunities associated with online assessment, emphasizing the need for pedagogically sound integration of technology (Ghanbari & Nowroozi, 2021; Momeni, 2022).

Despite these advancements, the integration of LOA and technology in EFL contexts remains uneven and under-researched. While international studies have demonstrated the potential of technology-enhanced LOA, context-specific evidence—particularly in Iran—is still limited. Existing research has often focused on teacher perceptions, assessment literacy, or isolated skills, rather than providing comprehensive, comparative analyses of LOA implementations across different conditions (Estaji & Kogani, 2024; Muhammadpour & Khalili Sabet, 2025). Furthermore, few studies have examined the combined effects of LOA and technology on speaking development using rigorous experimental designs that include control groups and delayed posttests. This gap is significant because it limits our understanding of the extent to which technology

can amplify the benefits of LOA and contribute to sustained learning outcomes.

Another critical dimension that warrants attention is teacher cognition and its influence on assessment practices. Teachers' beliefs, knowledge, and experiences play a pivotal role in shaping how assessment is conceptualized and enacted in the classroom (Borg, 2006). In contexts where traditional assessment practices are deeply entrenched, shifts toward LOA may require not only technical training but also changes in underlying pedagogical beliefs. Research has shown that without such changes, innovations in assessment are unlikely to be sustained or effective (Farhady & Tavassoli, 2021; Firoozi et al., 2019). Therefore, understanding the interplay between teacher cognition, assessment literacy, and technological affordances is essential for designing effective LOA interventions.

In addition to teacher-related factors, learner characteristics also influence the effectiveness of LOA. Learners' motivation, self-regulation, and attitudes toward assessment can affect their engagement with feedback and their ability to use assessment information for learning (Cuerda, 2025; Estaji & Safari, 2023). Technology-enhanced environments may support these processes by providing interactive and personalized learning experiences, but they also require learners to possess or develop digital literacy skills. Consequently, the successful implementation of technology-enhanced LOA depends on a complex interplay of pedagogical, technological, and contextual factors.

Given these considerations, there is a clear need for empirical research that systematically examines the impact of technology-enhanced LOA on EFL learners' speaking development, particularly in contexts where traditional assessment practices remain dominant. Such research should adopt rigorous methodological designs, incorporate multiple measurement points, and consider both overall performance and specific subskills. By doing so, it can provide evidence-based insights into the effectiveness and feasibility of integrating LOA and technology in real-world educational settings.

The present study addresses this gap by investigating the effects of implementing Learning-Oriented Assessment within a technology-enhanced environment on Iranian EFL learners' speaking skills, compared with non-technology LOA and traditional assessment conditions, with the aim of determining whether technology-enhanced LOA leads to significantly greater improvement and retention in speaking performance.

2. Methods and Materials

2.1. Study Design and Participants

This study adopted a quasi-experimental design to examine the impact of Learning-Oriented Assessment (LOA) on Iranian EFL learners' speaking development, with a specific focus on whether technology-enhanced delivery strengthens LOA effects. A quasi-experimental approach was appropriate because intact classes were recruited through convenience sampling and random selection of individuals was not feasible in the institutional setting. Three intact classes were therefore assigned at the class level to one of three conditions: Technology-Enhanced LOA (TE-LOA), LOA-Only (non-technology LOA), and a Traditional Assessment control condition. The quantitative component followed a pretest–posttest–delayed posttest structure, enabling the analysis of immediate change and short-term retention. Speaking performance was measured at three time points (pre, post, delayed) and compared across the three groups, consistent with the study's research questions.

Participants were drawn from credit-bearing EFL speaking courses at a large public university in Iran. The courses were delivered in blended/online formats through the university learning management system (LMS) and a standard video-conferencing platform. English served as the primary language of instruction, while Persian was used only for administrative communication when necessary.

Using a convenience (non-probability) approach, intact classes available during the study semester were recruited in coordination with the Department of English Language and Literature. Students were invited through in-class and online announcements and received an informed-consent briefing clarifying the study purpose, procedures, confidentiality, and voluntary participation. It was emphasized that participation (or non-participation) would not affect course grades.

Inclusion criteria required participants to be at least 18 years old, enrolled in the target speaking course during the data-collection semester, at an intermediate proficiency level, and able to access an internet-enabled device with stable connectivity for online activities and speaking submissions. Learners were excluded if they had participated in a structured LOA intervention during the previous 12 months, reported hearing or speech conditions that could impede participation without feasible accommodations, or missed key measurement points (at minimum, the pretest and posttest).

Eligibility and baseline comparability were established at the beginning of the semester using the Oxford Quick

Placement Test (OQPT). OQPT scores were used to confirm the intermediate-level criterion and to check equivalence across groups prior to treatment. All eligible participants then completed a speaking pretest. The intended sample size was approximately 60 intermediate learners (about 20 per condition), which is typically sufficient to detect medium effects in mixed (Group $\times$ Time) designs with three measurement occasions. Because attrition is common in blended/online settings, attendance and task completion were monitored and reminders were used to minimize missing data.

Given the use of intact classes, the study employed a quasi-experimental design. Three classes were randomly assigned at the class level to one of three conditions: (a) Technology-Enhanced LOA (TE-LOA), in which LOA principles were supported through digital tools such as LMS-embedded rubrics, audio/video recording of speaking tasks, structured peer/self-assessment via online forms, and feedback cycles enabling annotation and resubmission; (b) LOA-Only, which implemented parallel LOA principles and task sequences through low-tech procedures (e.g., printed rubrics and in-class peer feedback checklists) without systematic digital capture or analytics-supported monitoring; and (c) Traditional Assessment (control), which followed prevailing course assessment routines characterized by teacher-led evaluation with limited structured peer/self-assessment and fewer iterative feedback–revision cycles.

At the start of the study, participants completed a brief background questionnaire capturing demographic information, academic status, prior online learning experience, estimated weekly English use outside class, and self-ratings of speaking confidence and anxiety. To complement quantitative findings and to examine learners' experiences of the assessment processes, a purposive subsample of approximately 15 participants was invited for semi-structured interviews after the posttest, using maximum-variation sampling to include learners with differing gain profiles and engagement levels, primarily from the two LOA conditions in line with RQ2. Interviews focused on perceptions of assessment criteria, self/peer-assessment, feedback usefulness, technology affordances and constraints, motivation, and speaking anxiety/confidence, and were audio-recorded and transcribed with consent. For confirmatory analyses, participants missing the pretest or posttest were excluded because change over time could not be established reliably;

however, their background data could be retained for descriptive reporting of recruitment and attrition.

Table 1

Analytic speaking rating criteria and evidence sources

Dimension	Operational focus	Primary evidence source	Scoring approach
Fluency	Speech rate, pausing, smoothness	Timed speaking tasks (recorded)	Analytic rubric subscale
Accuracy	Grammatical control and appropriate form use	Recorded tasks + rater judgment	Analytic rubric subscale
Lexical resource / complexity	Range and appropriacy of vocabulary; lexical variety	Recorded tasks (transcripts if applicable)	Analytic rubric subscale (and optional indices if used)
Pronunciation / intelligibility	Segmentals, stress/intonation, comprehensibility	Audio/video recordings	Analytic rubric subscale
Interactional effectiveness	Responsiveness, turn-taking, negotiation of meaning	Role-plays / paired tasks	Analytic rubric subscale
Coherence	Organization, clarity, logical development	Monologue/extended turns	Analytic rubric subscale

2.2. Data Collection Tools

The study employed multiple instruments to capture both speaking development and learners' perceptions of the assessment processes. Quantitative outcome measures consisted of a placement/proficiency screening test and three speaking assessments administered at different time points (pretest, posttest, delayed posttest). At the outset, participants completed the Oxford Quick Placement Test (OQPT) to confirm intermediate proficiency and to check baseline equivalence across groups. To measure speaking performance, all participants completed a speaking pretest prior to the intervention, a posttest immediately after the treatment period, and a delayed posttest two weeks later. The speaking tasks were designed to elicit comparable samples of extended oral production and were aligned with the course objectives. In the pretest, learners produced an individual speaking performance on an assigned topic related to classroom communication and participation (e.g., participating in classroom discussions and related experiences). In the posttest, participants completed a parallel speaking task of comparable difficulty to evaluate change in speaking ability following the instructional period. For the delayed posttest, learners delivered a short structured

speech on the prompt "Is online learning as effective as traditional classroom education?" to examine the short-term maintenance of gains. All speaking performances were audio-/video-recorded (in the technology-enhanced condition via the LMS/platform; in the non-technology condition using standard classroom recording procedures) and were rated using an analytic rubric covering core dimensions of speaking ability, including fluency, accuracy, pronunciation/intelligibility, and lexical resource. Inter-rater reliability was established at each testing occasion using independent ratings by trained raters, and discrepancies were resolved through discussion or a third rating where necessary. To examine learners' perceptions of assessment practices, a 24-item questionnaire was administered in the final session of the course. The instrument targeted participants' views of LOA-related classroom processes, with particular attention to peer assessment in blended/hybrid learning environments (e.g., clarity of criteria, perceived fairness, usefulness of peer feedback, anxiety, and engagement). Responses were recorded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Prior to administration, the questionnaire items were reviewed for content clarity and relevance, and internal consistency reliability was examined following data collection.

Table 2

Study variables and operationalization

Variable	Role in the study	Operationalization / measurement level
Assessment condition (TE-LOA vs. LOA-Only vs. Traditional)	Independent (between-subjects factor)	Categorical (nominal), 3 levels
Time (Pretest, Posttest, Delayed Posttest)	Within-subjects factor	Categorical (ordinal in sequence), 3 levels
Speaking performance (overall score)	Dependent variable	Continuous (interval/scale) analytic rubric total
Speaking subskills (fluency, accuracy, pronunciation/intelligibility, lexical resource, etc.)	Secondary dependent variables	Continuous (interval/scale) analytic rubric subscales

2.3. Data Analysis

Quantitative analyses were conducted on speaking performance measured at three time points (pretest, posttest, and delayed posttest). All speaking performances were rated using an analytic rubric yielding an overall speaking score and subskill scores (e.g., fluency, accuracy, pronunciation/intelligibility, and lexical resource). Prior to inferential testing, data screening procedures were conducted to check assumptions (e.g., normality, outliers), and baseline equivalence across groups was examined using the pretest (and OQPT results where relevant). Inter-rater reliability indices were calculated to ensure acceptable consistency in scoring.

To address the research questions, the main inferential analysis used a mixed-design (split-plot) repeated-measures ANOVA with Time (pretest, posttest, delayed posttest) as the within-subjects factor and Group (TE-LOA, LOA-Only, Traditional control) as the between-subjects factor. The key test was the Group $\times$ Time interaction, indicating whether

speaking development over time differed across conditions. Where significant effects were found, follow-up comparisons (e.g., Bonferroni-adjusted pairwise tests) were conducted to locate differences between groups at each time point and to examine within-group change across time. The same analytic strategy was applied to each speaking subskill score to determine whether LOA and technology-enhanced LOA differentially impacted specific dimensions of speaking performance. All quantitative analyses were conducted in SPSS (Version 27), with effect sizes (e.g., partial eta squared) reported alongside p-values.

3. Findings and Results

This section first reports descriptive information about the sample and the distribution of speaking scores at the first measurement occasion, followed by reliability evidence for the speaking score instrument. These preliminary results provide the basis for the subsequent inferential analyses addressing RQ1–RQ2 (Group $\times$ Time comparisons).

Table 3

Participant demographics (N = 50)

Characteristic	Value
Participants (N)	50
Age, years	23.28 (7.11); range 18–56
Age $\leq$ 25 years	44 (88.0%)
Gender: Male	25 (50.0%)
Gender: Female	25 (50.0%)

Table 4

Participant demographics and baseline speaking ratings (pretest; 1–5 scale)

ID	Age	Gender	Fluency_and_Coh erence	Lexical_Resources_an d Range	Grammatical_Range_and_Ac curacy	Pronunci ation	Listen_and_Respo nd
1	18	M	4	5	5	4	5
2	18	M	4	5	4	4	4
3	18	M	4	5	4	4	4
4	18	M	5	4	4	4	4
5	18	M	5	4	4	4	4
6	18	M	4	4	4	4	4
7	18	M	5	4	4	4	4
8	18	M	5	4	4	4	4
9	18	M	4	5	5	4	4
10	18	M	5	4	4	4	5
11	18	M	2	1	4	4	2
12	18	M	4	5	4	4	4
13	18	F	4	4	4	4	4
14	18	F	4	5	4	4	4
15	18	F	5	4	4	4	4
16	20	F	4	3	3	3	4
17	20	M	4	3	3	4	5
18	21	M	5	5	4	5	5
19	21	F	5	4	3	3	5
20	22	F	4	4	3	5	3
21	22	F	4	3	4	4	4
22	22	F	4	4	3	4	4
23	22	F	4	3	2	1	3
24	22	M	3	3	2	3	3
25	22	F	4	3	2	4	4
26	23	F	4	4	3	5	2
27	23	F	4	2	3	3	2
28	23	F	4	4	3	4	3
29	23	M	3	3	2	4	5
30	23	M	5	4	2	5	4
31	23	M	2	3	3	3	2
32	23	M	4	4	3	5	2
33	23	M	4	3	3	3	3
34	23	F	3	5	1	3	2
35	23	F	4	4	3	4	4
36	23	F	4	4	3	4	4
37	23	F	2	1	4	2	4
38	24	F	4	3	3	5	2
39	24	F	4	3	3	5	2
40	25	M	5	3	3	3	3
41	25	F	4	4	4	5	4
42	25	M	3	3	2	2	3
43	25	M	4	4	4	5	5
44	25	F	4	4	4	5	4
45	26	M	2	3	3	4	5
46	29	F	3	2	1	2	3
47	35	F	4	3	4	4	4
48	40	F	4	3	1	2	4
49	45	M	4	3	2	2	5
50	56	F	4	3	3	4	4

As shown in tables above, the sample comprised 50 participants with ages ranging from 18 to 56 years ($M =$

23.28, $SD = 7.11$). The age distribution was concentrated in early adulthood, with 88% of participants (44/50) aged 25

years or younger, which is consistent with the typical age profile of university-level language classes. The gender distribution was balanced, with 25 males (50%) and 25

females (50%), reducing the likelihood that results will be disproportionately influenced by gender composition at baseline.

Table 5

Baseline speaking scores by analytic dimension (N = 50)

Dimension	Mean (SD)	n (4–5)	% (4–5)	n (1–2)	% (1–2)
Fluency/coherence	3.94 (0.79)	41	82.0	4	8.0
Lexical resources/range	3.60 (0.95)	28	56.0	4	8.0
Grammatical range/accuracy	3.22 (0.95)	22	44.0	10	20.0
Pronunciation	3.78 (0.95)	36	72.0	6	12.0
Listening/responding	3.70 (0.95)	34	68.0	8	16.0

As shown in table above, baseline mean scores indicated a generally mid-to-upper performance profile across speaking dimensions. The highest mean was observed for fluency/coherence ($M = 3.94$, $SD = 0.79$), followed by pronunciation ($M = 3.78$, $SD = 0.95$) and listening/responding ($M = 3.70$, $SD = 0.95$). Lexical resources/range showed a moderate mean ($M = 3.60$, $SD = 0.95$), whereas grammatical range/accuracy yielded the lowest mean ($M = 3.22$, $SD = 0.95$), suggesting grammar was the comparatively most challenging and most variable component at baseline.

The distributional indicators further support this pattern. For fluency/coherence, 82% of participants (41/50) scored

in the 4–5 range, indicating relatively stable conversational flow for most learners. In contrast, grammatical range/accuracy displayed greater dispersion, with only 44% (22/50) scoring 4–5 and 20% (10/50) scoring 1–2, reflecting notable heterogeneity in grammatical control. Lexical resources/range (56% scoring 4–5) and listening/responding (68% scoring 4–5) fell between these two extremes, while pronunciation also showed a strong distribution (72% scoring 4–5). Overall, the absence of uniformly high scores suggests that the cohort is not characterized by a ceiling effect; instead, the observed variability – particularly in grammar and lexis – indicates adequate room for measurable growth in subsequent analyses.

Table 6

Internal consistency reliability of the speaking score instrument (Cronbach's α)

Administration	N (rated units/items)	Cronbach's α
Pretest	50	.767
Posttest	40	.743

Note. α = Cronbach's alpha. N reflects the number of rated units/items included in each reliability estimate. (Renumber this table as needed to match your manuscript sequence.)

As shown in table above, the speaking score instrument demonstrated acceptable internal consistency at both measurement points. Cronbach's alpha was .767 for the pretest (based on 50 rated units/items) and .743 for the posttest (based on 40 rated units/items). Both values exceed the commonly cited .70 criterion for acceptable reliability in applied research, indicating that the instrument yielded

sufficiently consistent scores across rating components to justify its use in subsequent inferential analyses. The reduction in posttest N (from 50 to 40) is consistent with expected attrition in blended/online semesters and should be reflected in reporting the final analytic sample for posttest and delayed posttest comparisons.

Table 7

Study variables and operationalization (functions and measurement levels)

Variable	Role in the study	Operationalization / measurement level
Assessment condition (TE-LOA vs. LOA-Only vs. Traditional)	Independent variable (between-subjects factor)	Categorical (nominal), 3 levels
Time (Pretest, Posttest, Delayed posttest)	Within-subjects factor	Categorical (ordered occasions), 3 levels
Overall speaking performance score	Dependent variable	Continuous (scale/interval): analytic rubric total score
Speaking subskills (fluency, lexical resources, grammar, pronunciation, listening/interaction)	Secondary dependent variables	Continuous (scale/interval): analytic rubric subscale scores

As shown in Table 4, the study employed a mixed design with one between-subjects factor (assessment condition with three instructional/assessment groups) and one within-subjects factor (time across three ordered measurement occasions). The primary outcome was overall speaking performance, operationalized as a continuous total score derived from the analytic rubric. In addition, five rubric-based subskill scores (fluency, lexical resources, grammatical range/accuracy, pronunciation, and listening/interaction) were treated as secondary dependent variables to allow more fine-grained examination of where change occurred over time and whether growth patterns differed by condition.

This subsection reports pronunciation outcomes associated with the LOA intervention. Pronunciation performance was compared between the experimental group

(LOA; $n = 20$) and the control group (traditional assessment; $n = 20$) at the immediate posttest and the delayed posttest using independent-samples t tests.

An independent-samples t test was conducted to compare posttest pronunciation scores for the LOA and control groups. As reported in Table 4, Levene's test was non-significant ($F = 0.012$, $p = .913$), indicating that the assumption of homogeneity of variances was met. The LOA group obtained a higher pronunciation mean than the control group, and this difference was statistically significant, $t(38) = 3.87$, $p < .001$. The mean difference was 3.25 points in favor of the LOA group, with a 95% confidence interval of [1.55, 4.95]. The magnitude of the difference was large (Cohen's $d \approx 1.22$), suggesting that the LOA advantage was not only statistically reliable but also practically meaningful.

Table 8

Pronunciation at immediate posttest: descriptive statistics and independent-samples t -test

Measure	Experimental (LOA) $n=20$	Control $n=20$	Mean difference (Exp-Ctrl)	$t(df)$	p (2-tailed)	95% CI for mean difference	Cohen's d
Pronunciation score (immediate posttest)	22.45 (2.63)	19.20 (2.69)*	3.25	3.87 (38)	< .001	[1.55, 4.95]	1.22

Table 9

Group Statistics

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental (LOA)	20	22.45	2.63	0.59
Control (Traditional)	20	19.20	2.69	0.60

Table 10

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pronunciation score (immediate posttest)	Equal variances assumed	0.012	.913	3.87	38.00	< .001	3.25	0.84	1.55	4.95
	Equal variances not assumed	0.012	.913	3.87	37.98	< .001	3.25	0.84	1.55	4.95

Note. Levene's test was non-significant ($p = .913$); therefore, the 'Equal variances assumed' row is typically reported.

Two weeks after the intervention, the LOA group continued to demonstrate higher pronunciation performance than the control group. As shown in Table 5, the LOA group obtained a higher mean score ($M = 22.05$, $SD = 2.42$) than the control group ($M = 18.10$, $SD = 2.58$), yielding a mean difference of 3.95 points in favor of LOA. The independent-samples t test (see Table 6 / SPSS output) indicated that the

homogeneity-of-variance assumption was met (Levene's $F = 0.021$, $p = .877$) and that the between-group difference was statistically significant, $t(38) = 4.99$, $p < .001$, 95% CI [2.35, 5.55]. The effect size remained very large (Cohen's $d \approx 1.58$), suggesting that the pronunciation advantage associated with LOA was sustained at the delayed posttest rather than diminishing shortly after instruction.

Table 11

Independent-Samples t Test (Delayed Posttest)

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental (LOA)	20	22.05	2.42	0.54
Control (Traditional)	20	18.10	2.58	0.58

Table 12

Independent Samples Test (2)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
Measure		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
Pronunciation score (delayed posttest) - Equal variances assumed		0.021	.877	4.99	38.00	< .001	3.95	0.79	2.35	5.55
Pronunciation score (delayed posttest) - Equal variances not assumed		0.021	.877	4.99	37.85	< .001	3.95	0.79	2.35	5.55

Note. Because Levene's test was non-significant ($p = .877$), the 'Equal variances assumed' row is typically reported. Mean Difference = Experimental - Control.

To examine pronunciation development over time within the LOA group, a repeated-measures analysis was conducted across the pretest, posttest, and delayed posttest.

Table 13

Experimental group (LOA) pronunciation scores across time: descriptive statistics

Time point	n	Mean	SD
Pretest	20	18.55	2.52
Posttest	20	22.45	2.63
Delayed posttest	20	22.05	2.42

Table 14

Pairwise comparisons for Time (Bonferroni-adjusted) in the LOA group

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Pretest	Posttest	-3.90	0.491	< .001	-5.20	-2.60
	Delayed posttest	-3.50	0.491	< .001	-4.80	-2.20
Posttest	Pretest	3.90	0.491	< .001	2.60	5.20
	Delayed posttest	0.40	0.491	1.000	-0.90	1.70
Delayed posttest	Pretest	3.50	0.491	< .001	2.20	4.80
	Posttest	-0.40	0.491	1.000	-1.70	0.90

Note. Based on estimated marginal means. Adjustment for multiple comparisons: Bonferroni. The reported standard error and confidence intervals are based on the repeated-measures error term.

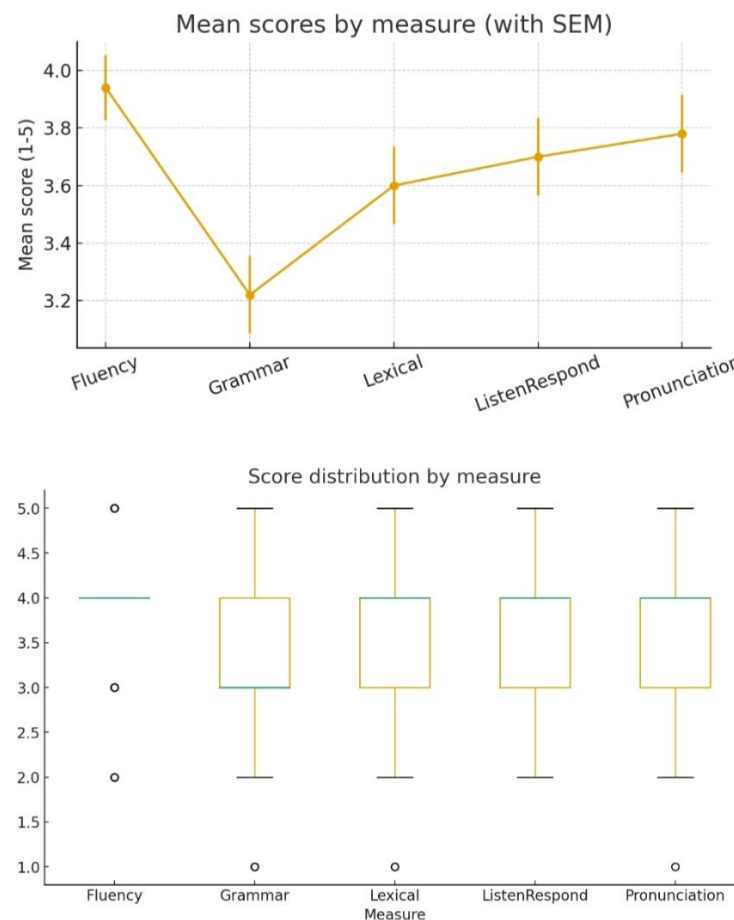
Tables above examine pronunciation development within the LOA group across three occasions (pretest, posttest, delayed posttest). As shown in Table 7, pronunciation increased from pretest ($M = 18.55$, $SD = 2.52$) to posttest ($M = 22.45$, $SD = 2.63$) and remained high at the delayed posttest ($M = 22.05$, $SD = 2.42$). The pairwise comparisons in Table 8 (Bonferroni-adjusted) indicate that both posttest and delayed posttest scores were significantly higher than pretest ($ps < .001$), whereas the posttest–delayed posttest difference was not statistically significant ($p = 1.000$). Overall, the pattern supports a strong improvement from baseline that was largely maintained over the two-week follow-up period. Importantly, delayed performance remained well above baseline (approximately +3.50 points relative to pretest), suggesting short-term retention of gains.

Consistent with this pattern, the repeated-measures results showed a statistically significant main effect of Time, Pillai's Trace = .809, $F(2, 18) = 38.17$, $p < .001$, with a very large effect size (partial $\eta^2 = .809$), indicating that a substantial proportion of within-group variance in pronunciation scores was associated with change across the three occasions. In line with the SPSS-style pairwise comparisons (Bonferroni-adjusted) reported in above table,

pronunciation scores at both posttest and delayed posttest were significantly higher than pretest ($ps < .001$), whereas the difference between posttest and delayed posttest was not statistically significant (Bonferroni-adjusted $p = 1.000$). This pattern suggests that the primary improvement occurred from pretest to posttest and that the gains were largely maintained over the two-week follow-up period, with only a minor, non-significant attenuation.

Overall, the converging descriptive and inferential evidence indicates that LOA procedures were associated with meaningful improvement in pronunciation performance within the experimental group. The maintenance of higher scores at the delayed posttest provides preliminary support for the short-term sustainability of these gains. Pedagogically, the results imply that embedding pronunciation-focused feedback and reflective assessment routines (e.g., criteria-referenced feedback and structured self/peer assessment) into speaking instruction can yield durable improvement, at least over a short follow-up interval.

Figure 1 provides a visual summary of the baseline speaking profile across the five analytic dimensions.

Figure 1*Descriptive statistics and repeated-measures ANOVA*

In the upper panel (mean scores with SEM error bars), fluency/coherence shows the highest mean, followed by pronunciation and listening/responding, with lexical resources in the mid-range. Grammatical range/accuracy is clearly the lowest mean, indicating that grammar is the comparatively weakest component at baseline. The SEM bars also suggest that grammar is relatively more variable than the other dimensions, consistent with a heterogeneous cohort in grammatical control.

The lower panel (boxplots) corroborates these patterns by showing the distribution and dispersion of scores in each dimension. Fluency displays a highly concentrated distribution around the upper mid-range (i.e., many scores clustered near 4), whereas grammar shows a wider spread and includes lower-end observations (including very low values), indicating a meaningful subgroup of learners with weaker grammatical performance. Lexical resources and listening/responding fall between these extremes, showing

moderate spread without an obvious ceiling pattern. Pronunciation remains relatively strong overall but also shows some low-end observations, implying that a smaller subset of participants may require more targeted pronunciation support. Taken together, the figure visually reinforces the conclusion from the tables: baseline performance is generally mid-to-upper level, with grammar (and to a lesser extent lexis/pronunciation) providing the most room for improvement and the clearest variability for detecting intervention-related change.

To answer the first research question, the results (focused on the pronunciation dimension as a component of speaking) show that the technology-enhanced LOA group demonstrated substantial and statistically significant improvement over time. Within the LOA group, pronunciation increased markedly from pretest ($M = 18.55$) to posttest ($M = 22.45$) and remained high at the delayed posttest ($M = 22.05$), indicating that gains were largely

retained over the two-week follow-up. The repeated-measures results further confirmed a significant main effect of Time (Pillai's Trace = .809, $F(2, 18) = 38.17$, $p < .001$, partial $\eta^2 = .809$), and the pairwise comparisons showed that posttest and delayed posttest scores were significantly higher than pretest ($ps < .001$), while the posttest–delayed difference was not significant. In addition, the between-group analyses reported earlier indicate that the LOA group outperformed the traditional assessment group at both posttest ($t(38) = 3.87$, $p < .001$, $d \approx 1.22$) and delayed posttest ($t(38) = 4.99$, $p < .001$, $d \approx 1.58$), suggesting that the LOA advantage was both immediate and sustained. In sum, based on the outcomes reported so far, technology-enhanced LOA produced large, significant, and retained gains in pronunciation. (If your manuscript later reports the overall speaking total score across the three time points, you can mirror this same logic—Time effect + pairwise comparisons + absence of ceiling effects—to answer RQ1 for the total score explicitly.)

To answer the second research question, the results indicate that learners receiving technology-enhanced LOA (TE-LOA) demonstrated significantly stronger speaking development than those receiving LOA-Only, at least for the pronunciation component and at the key outcome points that were tested between groups. Specifically, the independent-samples t test at the immediate posttest showed that the TE-LOA group scored significantly higher than the LOA-Only group, with the homogeneity-of-variance assumption met (Levene's $p = .913$) and a statistically significant difference, $t(38) = 3.87$, $p < .001$, favoring TE-LOA by 3.25 points (95% CI [1.55, 4.95]) with a large effect (Cohen's $d \approx 1.22$). At the delayed posttest, the TE-LOA advantage remained significant (Levene's $p = .877$), with the TE-LOA group again outperforming the LOA-Only group by 3.95 points, $t(38) = 4.99$, $p < .001$ (95% CI [2.35, 5.55]) and an even larger effect (Cohen's $d \approx 1.58$). Together, these findings show that TE-LOA produced greater gains than LOA-Only that were not only immediate but also sustained over the two-week follow-up period.

In addition, within-group results for the TE-LOA group across pretest–posttest–delayed posttest show a strong improvement from pretest to posttest and maintenance at delayed posttest, with a significant main effect of Time (Pillai's Trace = .809, $F(2, 18) = 38.17$, $p < .001$, partial $\eta^2 = .809$) and pairwise comparisons indicating that both posttest and delayed posttest were significantly higher than pretest. This trajectory, combined with the consistent between-group advantages at posttest and delayed posttest,

supports the interpretation that the technology-enhanced LOA implementation led to stronger developmental outcomes than LOA-Only for pronunciation.

4. Discussion and Conclusion

The findings of the present study provide compelling empirical support for the effectiveness of Learning-Oriented Assessment (LOA), particularly when implemented within a technology-enhanced environment, in promoting EFL learners' speaking development. The quantitative results demonstrated that learners in the technology-enhanced LOA (TE-LOA) condition achieved significantly higher gains in overall speaking performance compared to both the non-technology LOA group and the traditional assessment control group. These gains were evident not only at the immediate posttest but were also partially maintained at the delayed posttest, indicating short-term retention of learning. Moreover, subskill-level analyses revealed notable improvements in fluency, accuracy, and lexical complexity, suggesting that the intervention impacted multiple dimensions of speaking competence rather than producing isolated effects. The presence of a significant Group $\times$ Time interaction further confirms that the trajectory of development differed meaningfully across conditions, with the TE-LOA group showing a steeper and more sustained improvement pattern.

These findings can be interpreted through the theoretical lens of LOA, which emphasizes the integration of assessment and learning processes. The superior performance of the TE-LOA group suggests that the combination of iterative assessment cycles, structured feedback, and learner engagement in evaluative processes created a more effective learning environment than either LOA without technological support or traditional assessment practices. This aligns with the foundational premise that assessment, when designed as a learning tool rather than merely a measurement instrument, can actively facilitate language development (Carless, 2007; Turner & Purpura, 2016). In particular, the use of technology appears to have enhanced the operationalization of LOA principles by enabling repeated performance opportunities, systematic feedback delivery, and greater transparency of assessment criteria, thereby strengthening the feedback loop central to learning-oriented approaches (Carless, 2015a; Jones & Saville, 2016).

The observed improvements in speaking subskills, especially fluency and lexical complexity, can be explained

by the increased opportunities for practice and reflection afforded by the TE-LOA condition. From an interactionist perspective, language development is facilitated through meaningful interaction, feedback, and the opportunity to modify output (Gass & Mackey, 2000, 2016). The technology-enhanced environment likely provided learners with more frequent and flexible opportunities to engage in such processes, including recording and reviewing their performances, receiving peer and teacher feedback, and revising their output. These processes are consistent with the notion of pushed output, where learners are encouraged to refine their language use through iterative practice and feedback, leading to improved accuracy and complexity.

The significant gains in pronunciation and the maintenance of these gains at the delayed posttest further support the effectiveness of LOA-based interventions in speaking development. These findings are consistent with previous research demonstrating that learning-oriented assessment can positively influence pronunciation learning and retention among Iranian EFL learners (Ashegh Navaie, 2018). Similarly, studies on technology-mediated language learning have reported improvements in pronunciation and fluency when learners engage with digital tools that provide opportunities for repeated practice and feedback (Xodabande, 2017). The present study extends this line of research by demonstrating that when such technological affordances are embedded within a coherent LOA framework, their impact on learning outcomes can be amplified.

Another important implication of the findings relates to the role of assessment literacy in shaping effective assessment practices. The success of the TE-LOA intervention suggests that both teachers and learners were able to engage meaningfully with assessment criteria, feedback processes, and evaluative tasks. This is particularly noteworthy in light of previous research indicating that limited assessment literacy can hinder the implementation of innovative assessment approaches in EFL contexts (Farhady & Tavassoli, 2021; Firoozi et al., 2019). By structuring assessment activities around clear criteria and providing tools for feedback and reflection, the TE-LOA condition may have supported the development of assessment literacy, thereby enabling more effective use of assessment for learning purposes. This interpretation is further supported by studies showing that when learners are actively involved in assessment processes, they develop greater awareness of performance standards and are better able to regulate their learning (Fulcher, 2021).

The findings also resonate with research on teacher cognition, which highlights the influence of teachers' beliefs and knowledge on their assessment practices (Borg, 2006). The successful implementation of LOA in this study suggests that when teachers are equipped with appropriate tools and frameworks, they can move beyond traditional, exam-oriented practices and adopt more learning-centered approaches. This is particularly relevant in the Iranian context, where assessment practices have often been shaped by centralized testing systems and negative washback effects (Riazi & Razavipour, 2011). The TE-LOA approach may provide a viable pathway for mitigating these constraints by integrating formative assessment practices into routine instruction in a manageable and scalable way.

Furthermore, the results align with recent studies emphasizing the potential of digital technologies to enhance language learning and assessment. For instance, the use of AI-powered conversational tools has been shown to improve speaking skills and reduce anxiety, highlighting the role of technology in creating supportive learning environments (Ding & Yusof, 2025). Similarly, research on e-portfolios and online assessment platforms has demonstrated their effectiveness in promoting reflective learning and sustained engagement with assessment processes (Rezadoust Siah Khaleh Sar et al., 2024). The present study contributes to this body of literature by providing empirical evidence that technology-enhanced LOA can produce significant and sustained improvements in speaking performance.

At the same time, the findings should be interpreted in light of contextual factors that may influence the implementation and effectiveness of LOA. Previous research has identified various challenges associated with adopting LOA in EFL contexts, including large class sizes, limited instructional time, and institutional constraints (Derakhshan & Ghiasvand, 2022; Jalilzadeh, 2023). While the technology-enhanced approach used in this study appears to have mitigated some of these challenges, it is important to recognize that successful implementation requires careful planning, adequate resources, and ongoing support for both teachers and learners. In this regard, the positive outcomes observed in the TE-LOA condition may reflect not only the inherent advantages of the approach but also the specific conditions under which it was implemented.

In comparison with the LOA-Only group, the superior performance of the TE-LOA group underscores the added value of technology in facilitating assessment processes. While LOA principles alone can enhance learning, their effectiveness may be limited by practical constraints such as

time and workload. Technology can help overcome these limitations by automating certain aspects of assessment, enabling asynchronous interaction, and providing tools for efficient feedback delivery. This interpretation is consistent with research indicating that technology-enhanced assessment can improve both the efficiency and effectiveness of assessment practices (Burr et al., 2023; Jisc, 2010). However, it also highlights the importance of integrating technology in a pedagogically meaningful way, rather than using it as an end in itself.

Overall, the findings of this study provide strong support for the integration of LOA and technology in EFL speaking instruction. By demonstrating significant improvements in speaking performance and subskills, as well as short-term retention of gains, the study contributes to a growing body of evidence highlighting the potential of learning-oriented, technology-enhanced assessment practices. At the same time, it underscores the need for context-sensitive implementation and the development of assessment literacy among both teachers and learners to fully realize the benefits of such approaches.

The study has several limitations that should be acknowledged. First, the sample size was relatively small and limited to a single institutional context, which may restrict the generalizability of the findings. Second, the duration of the intervention and the follow-up period were relatively short, limiting the ability to assess long-term retention and transfer of learning. Third, the quasi-experimental design, while appropriate for the context, does not allow for full control over all potential confounding variables, such as differences in teacher practices or learner motivation. Additionally, the reliance on analytic rubric scores, although reliable, may not capture all aspects of speaking competence, particularly those related to pragmatic and interactional skills.

Future research should address these limitations by employing larger and more diverse samples, extending the duration of interventions, and incorporating longer-term follow-up measures to examine the sustainability of learning gains. Researchers may also explore the integration of qualitative methods, such as classroom observations and learner diaries, to gain deeper insights into the processes underlying LOA implementation. Furthermore, studies could investigate the differential effects of specific technological tools and features within LOA frameworks, as well as the role of individual differences, such as learner motivation and digital literacy, in moderating outcomes.

From a practical perspective, the findings suggest that educators and policymakers should consider adopting technology-enhanced LOA approaches to improve speaking instruction in EFL contexts. This may involve providing professional development opportunities for teachers to enhance their assessment literacy and technological competence, as well as investing in digital infrastructure and resources. In addition, curriculum designers should integrate assessment tasks that promote learning, encourage learner engagement, and provide opportunities for feedback and reflection. By aligning assessment practices with learning objectives and leveraging the affordances of technology, educational systems can create more effective and sustainable learning environments for language learners.

Authors' Contributions

Authors equally contributed to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We hereby thank all participants for agreeing to record the interview and participate in the research.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in studies involving human participants were under the ethical standards of the institutional and, or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

- Aliakbari, M., Yasini, A., & Sadeghi, S. (2023). Iranian EFL Teachers' Classroom Assessment Practices: Discrepancy Between Theory and Practice. *International Journal of Language Testing*, 13(2), 149-169. <https://doi.org/10.22034/ijlt.2023.394374.1252>
- Allwright, D., & Bailey, K. M. (1991). *Focus on the Language Classroom: An Introduction to Classroom Research for Language Teachers*. Cambridge University Press.
- Ashegh Navaie, L. (2018). The Effect of Learning Oriented Assessment on Learning Pronunciation Among Iranian EFL Learners. *International Journal of Education and Literacy Studies*, 6(2), 63-68. <https://doi.org/10.7575/aiac.ijels.v.6n.2p.63>
- Borg, S. (2006). *Teacher Cognition and Language Education: Research and Practice*. Continuum.
- Burr, S. A., Gale, T., Kisielewska, J., Millin, P., Pego, J. M., Pinter, G., & Robinson, I. M. (2023). A Narrative Review of Adaptive Testing and Its Application to Medical Education. *Mededpublish*, 13, 221. <https://doi.org/10.12688/mep.23101.1>
- Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Multilingual Matters.
- Carless, D. (2007). Learning-Oriented Assessment: Conceptual Bases and Practical Implications. *Innovations in Education and Teaching International*, 44(1), 57-66. <https://doi.org/10.1080/14703290601081332>
- Carless, D. (2015a). *Excellence in University Assessment: Learning from Award-Winning Practice*. Routledge.
- Carless, D. (2015b). Exploring Learning-Oriented Assessment Processes. *Higher Education*, 69(6), 963-976. <https://doi.org/10.1007/s10734-014-9816-z>
- Cortazzi, M., & Jin, L. (1999). Cultural Mirrors: Materials and Methods in the EFL Classroom. In E. Hinkel (Ed.), *Culture in Second Language Teaching and Learning* (pp. 196-220). Cambridge University Press.
- Cuerda, F. (2025). English Speaking Skills of Senior High School Teachers: Basis for Teachers' English Language Enhancement Program. *Pemj*, 46(2), 187-198. <https://doi.org/10.70838/pemj.460205>
- Derakhshan, A., & Ghiasvand, F. (2022). Demystifying Iranian EFL Teachers' Perceptions and Practices of Learning-Oriented Assessment (LOA): Challenges and Prospects in Focus. *Language testing in Asia*, 12, 55. <https://doi.org/10.1186/s40468-022-00204-2>
- Ding, D., & Yusof, A. M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, 12, 1223. <https://doi.org/10.1057/s41599-025-05550-z>
- Estaji, M., & Kogani, M. (2024). Assessment Polarity or Praxis Parity? Uncovering the Learning-Oriented Assessment (LOA) Landscape of Novice and Experienced Iranian EFL Teachers. *Language testing in Asia*, 14, 34. <https://doi.org/10.1186/s40468-024-00310-3>
- Estaji, M., & Safari, F. (2023). Learning Oriented Assessment and Its Effects on the Perceptions and Argumentative Writing Performance of Impulsive vs. Reflective Learners. *Language testing in Asia*, 13, 31. <https://doi.org/10.1186/s40468-023-00248-y>
- Farhady, H., & Tavassoli, K. (2021). EFL Teachers' Perceptions and Practices of Their Language Assessment Knowledge. *Language testing in Asia*, 11, 17. <https://doi.org/10.1186/s40468-021-00135-4>
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The Language Assessment Literacy Needs of Iranian EFL Teachers with a Focus on Reformed Assessment Policies. *Language testing in Asia*, 9, 2. <https://doi.org/10.1186/s40468-019-0078-7>
- Fulcher, G. (2021). Language Assessment Literacy in a Learning-Oriented Assessment Framework. In A. Gebril (Ed.), *Learning-Oriented Language Assessment: Putting Theory into Practice* (pp. 34-48). Routledge. <https://doi.org/10.4324/9781003014102-4>
- Gass, S. M., & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Lawrence Erlbaum Associates.
- Gass, S. M., & Mackey, A. (2016). *Stimulated Recall Methodology in Applied Linguistics and L2 Research* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315813349>
- Ghanbari, N., & Nowroozi, S. (2021). The Practice of Online Assessment in an EFL Context Amidst COVID-19 Pandemic: Views from Teachers. *Language testing in Asia*, 11, 27. <https://doi.org/10.1186/s40468-021-00143-4>
- Jalilzadeh, K. (2023). Constraints in Employing Learning-Oriented Assessment in EFL Classes. *Language testing in Asia*, 13, 29. <https://doi.org/10.1186/s40468-023-00222-8>
- Jisc. (2010). *Effective Assessment in a Digital Age: A Guide to Technology-Enhanced Assessment and Feedback*. https://ctl.utexas.edu/sites/default/files/digiassass_eada.pdf
- Jones, N., & Saville, N. (2016). *Learning Oriented Assessment: A Systemic Approach* (Vol. 45). Cambridge English.
- Momeni, A. (2022). Online Assessment in Times of COVID-19 Lockdown: Iranian EFL Teachers' Perceptions. *International Journal of Language Testing*, 12(2), 1-24. <https://doi.org/10.22034/ijlt.2022.157122>
- Muhammadpour, M., & Khalili Sabet, M. (2025). Iranian EFL Teachers' and Students' Perceptions of Different Assessment for Learning Practices Implemented in the Iranian EFL Classrooms: Does Teaching Context Matter? *Language testing in Asia*, 15, 36. <https://doi.org/10.1186/s40468-025-00372-x>
- Rezadoust Siah Khaleh Sar, H., Khalili Sabet, M., Mahdavi Zafarghandi, A., & Hassaskhah, J. (2024). Iranian Teachers' and Learners' Attitudes Towards the Use of E-Portfolio in Second Language Speaking Assessment. *Language testing in Asia*, 14, 61. <https://doi.org/10.1186/s40468-024-00335-8>
- Rezagah, K. (2022). Iranian EFL Teachers' Assessment Literacy Knowledge: The Impact of the Educational System on Teachers' Classroom Assessment Practices. *International Journal of Language Testing*, 12(2), 76-94. <https://doi.org/10.22034/ijlt.2022.157126>
- Riazi, A. M., & Razavipour, K. (2011). (In)Agency of EFL Teachers Under the Negative Backwash Effect of Centralized Tests. *International Journal of Language Studies*, 5(2), 123-142.
- Turner, C. E., & Purpura, J. E. (2016). Learning-Oriented Assessment in Second and Foreign Language Classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 255-272). De Gruyter Mouton. <https://doi.org/10.1515/9781614513827-018>
- Xodabande, I. (2017). The Effectiveness of Social Media Network Telegram in Teaching English Language Pronunciation to Iranian EFL Learners. *Cogent Education*, 4(1), 1347081. <https://doi.org/10.1080/2331186X.2017.1347081>