




Research Article

The Impact of Learning-Oriented Assessment on Iranian Intermediate EFL Learners' Reading Comprehension and Strategy Use: A Mixed-Methods Study

Pegah Doroudi¹ , Mohammad Taghi Farvardin² ✉, Ghafour Rezaie Golandouz³ 

¹ Department of Foreign Languages Teaching, SR.C., Islamic Azad University, Tehran, Iran

² Department of Foreign Languages Teaching, Ahv.C., Islamic Azad University, Ahvaz, Iran

³ Department of Foreign Languages Teaching, Ga.C., Islamic Azad University, Garmsar, Iran

Abstract

Learner-oriented assessment (LOA) aims to combine assessment and teaching using feedback, self- and peer assessment, and reflective learning. Despite LOA's well-theorized basis, its specific impacts in real classrooms have remained underexplored. This research used a sequential exploratory mixed-method design to explore the effect of LOA on the Iranian intermediate English as a Foreign Language (EFL) learners' reading comprehension and reading strategy use. The experimental and control groups were drawn from two intact institute classes. In a 10-week intervention that consisted of 20 sessions (two sessions per week), the experimental group participated in rubric-guided self- and peer assessment exercises with formative feedback, while the control group received conventional summative tests. Quantitatively, reading comprehension was found to have significantly increased among the LOA group, and this gain was higher when contrasted to smaller changes in the control group. A one-way MANOVA, with group as the between-subjects factor and time as the within-subjects factor, was also performed on the SORS three subscales, which also yielded a significant group \times time interaction, all with large posttest effects in favor of the LOA group on overall, problem-solving, and support strategies. Qualitative data from semi-structured interviews corroborated the quantitative findings, highlighting how LOA promoted learners' autonomy, metacognitive awareness, and collaboration in problem-solving by making links to sociocultural theory. Overall, the results indicate that the LOA could develop students' reading comprehension and strategy use in EFL classrooms.

Keywords: formative feedback, learning-oriented assessment, metacognitive awareness, mixed-methods, reading comprehension, reading strategy use, sociocultural theory

Cite as: Doroudi, P., Farvardin, M.T., & Rezaie Golandouz, Gh. (2025). The impact of learning-oriented assessment on Iranian intermediate EFL learners' reading comprehension and strategy use: A mixed-methods study. *Mixed-Methods Studies in English Language Teaching*, 2(2), 79-100. <https://doi.org/10.71873/mslt.2025.1213918>

Received: 03/08/2025; **Revised:** 05/09/2025; **Accepted:** 08/09/2025; **Published:** 16/09/2025

© The Author(s), 2025 ✉ mtfarvardin@iau.ac.ir Publisher: Qom Islamic Azad University



1. Introduction

With regard to English as a foreign language (EFL) classroom teaching, conventional assessment activities have been centered on summative testing, such as national examinations that mainly measure the outcome but not learning progress (Carless, 2012; Jones & Saville, 2016). These methods threaten to promote a form of rote learning at the expense of an interactive, developmental four-language learning process (Purpura, 2016). In contrast, learning-oriented assessment (LOA) has been recommended as an approach that binds assessment and instruction, privileging formative feedback, metacognitive contemplation, and cooperative learning (Nurjain et al., 2023).

LOA is rooted in sociocultural theory, treating assessment as a means for mediating learning through scaffolding and social engagement (Vygotsky, 1978). In this context, dynamic assessment (DA) integrates teaching with testing so that teachers can offer individualized assistance that promotes students' development (Carless, 2015). Recent research in Iran has shown that LOA activities can enhance learners' metacognitive awareness and autonomy (Derakhshan & Ghiasvand, 2022; Ghaneiarani et al., 2024; Khalili et al., 2024). However, reading comprehension has been relatively less investigated (Ghaneiarani et al., 2024). The conventional reading instruction in Iran tends to emphasize the product-based testing rather than explicit strategy instruction, thus making it difficult for the students to practice and internalize efficient reading strategies (Khalili et al., 2024). LOA, which integrates strategy generation in the very process of task-solving, is a potential way to compensate for these deficiencies (Beikmohammadi et al., 2020).

While positive effects of LOA have been found in different settings around the world, the body of research focusing on the use of LOA in the Iranian classroom for reading comprehension and reading strategy use is still quite limited (Beikmohammadi et al., 2020). However, the perception of LOA and how scaffolded assessment strategies may promote strategy transfer across genres or texts is an area that has not been addressed in current research (Derakhshan & Ghiasvand, 2022; Jalilzadeh & Coombe, 2023). In addition, the demographic antecedents of the effectiveness of LOAs (e.g., gender, proficiency level) have been overlooked to a large extent.

The present study is significant as it is an attempt to broaden LOA research in reading comprehension and reading strategy use to the Iranian EFL classes, where it has obvious pedagogical and theoretical implications. By

integrating quantitative metrics with qualitative observations, the present study attempts to capture not only if LOA works to enhance reading performance but also how learners react to and experience the practices. Such results may illuminate curriculum design, teacher training, and classroom implementation of LOA in settings where summative testing presides. According to the above rationale, the research could be formulated in the following questions:

RQ1: Does learning-oriented assessment significantly affect EFL learners' reading comprehension?

RQ2: Does learning-oriented assessment significantly affect EFL learners' reading strategy use?

RQ3: What are EFL learners' perceptions of learning-oriented assessment?

2. Literature Review

LOA is based on sociocultural theory (Vygotsky, 1978), which considers that learning is a social process facilitated by scaffolding and interaction with more expert others. DA integrates these principles through the combination of instruction and assessment, which allows teachers to adjust the amount of scaffold support to meet the students' needs (Brown, 2018). However, there are several practical obstacles. For example, DA presumes small classes and close teacher-student interaction that are sometimes lacking in EFL settings (Carless, 2015). Similarly, although self- and peer assessment rely on metacognition, studies question the dependability and validity of self- and peer assessment, particularly with low-proficiency students (Al-Abri et al., 2025; Black & Wiliam, 1998; Yan & Carless, 2022). Therefore, though LOA is conceptually convincing, its classroom implementation requires caution.

Several paradigms try to bridge the gap between LOA and practice. Jones and Saville's (2016) classroom within an LOA model has learning and assessment as irremovable elements, where class activities lie at the center of the cycle, serving both learning and assessment purposes. This model is based on five principles: aligning activities with curriculum, designing tasks to serve as a means of learning and to capture learning, using explicit criteria, setting up feedback loops, and addressing the affective dimension of assessment to sustain attention (Jones & Saville, 2016). These principles combined establish a task design, engagement, feedback, and adaptation cycle that puts learning at the center. In this context, self- and peer assessment have particular merits in that, with well-defined criteria, they can promote metacognitive learning and learner autonomy (Yan & Carless, 2022).

For EFL reading, stimulating comprehension involves not only higher-level cognitive abilities (e.g., inferencing) but also language-specific strategies (e.g., skimming, scanning, and self-monitoring) (Grabe, 2009; Zhang, 2001). LOA can inform strategy construction by incorporating reflection and feedback into reading tasks. For example, reflective journals and rubric-guided peer assessment engage students in the description of reasoning and self-regulatory activities (Beikmohammadi et al., 2020). The evidence in practice is mixed, though. Banitalebi and Ghiasvand (2023) demonstrated that scaffolded questioning enhanced learners' self-monitoring but did not exert a significant impact on the test scores, underscoring tensions between formative practices and exam-oriented settings. Ghaneiarani et al. (2024) also reported an increase in writing through LOA-based feedback. Most of these studies are promising; however, there are some methodological concerns, such as small sample sizes, short interventions, and lack of control groups.

LOA's effectiveness is also mediated by teachers' and learners' attitudes. Nurjamin et al. (2023) found that LOA decreased anxiety levels and promoted reflective thinking in learners, particularly when the assessment was perceived as supportive reckoning rather than retributive. In the Iranian context, Derakhshan and Ghiasvand (2022) identified the teachers' attitudes toward LOA, noting that while they expressed positive attitudes, certain institutional constraints and lack of training were at play. Similar barriers were recognized by Jalilzadeh and Coombe (2023), such as prescriptive curricula and inadequate support by administration. Khalili et al. (2024) demonstrated the Iranian teachers' institutional readiness for LOA provided that they are trained, but they called for practical instruments and systemic modifications to foster sustainability.

Nevertheless, many open questions still remain in spite of these works. One reason is that the majority of LOA research in Iran has been conducted on writing or speaking without much attention toward reading comprehension or the use of strategies. Second, the methodological quality is mixed, with many studies based on small, short-term samples and self-reported perceptions (Banitalebi & Ghiasvand, 2023; Jalilzadeh & Coombe, 2023). Thirdly, theoretical models, such as Carless' (2015) model, are often referenced yet seldom questioned for their suitability in EFL learning environments, in which large classes and exam-based curricula are widely found. Finally, although metacognition is a cornerstone of LOA, there is scant empirical evidence about how students' ability is developed in terms of reflection and evaluation over time.

To summarize, the literature indicates that while there is considerable theoretical and practical interest in LOA, there are also notable areas where LOA has remained unexplored. Little is known about the effects of LOA on EFL learners' reading comprehension and the use of learning strategies with meager attention to exam-oriented contexts such as Iran. The present study fills this gap by juxtaposing quantitative and qualitative data on the impact of LOA in Iranian EFL reading classrooms.

3. Method

3.1. Design

A sequential mixed-methods experimental design (Creswell & Creswell, 2022) was implemented, incorporating a quasi-experimental quantitative phase followed by a qualitative phase to address the three research questions. This design made it possible to assess the effects of the intervention on reading comprehension and reading strategy use, and then to explore students' perceptions in more depth in order to account for processes of change. This study involved whole classes because of institutional scheduling constraints, and randomization occurred at the class level rather than the individual level. Quantitative data comprised pre- and posttests of reading comprehension and survey of reading strategies (SORS; Mokhtari & Sheorey, 2002) scores. For the qualitative phase, semi-structured interviews with 10 experimental participants were conducted, audio-recorded, transcribed, and thematically analyzed.

3.2. Participants

An a priori power analysis using G*Power 3.1 (Faul et al., 2009) indicated that at least 52 participants were needed in the total sample to detect a medium effect size (Cohen's $d = 0.50$) with $\alpha = .05$ and power = .80. For the MANOVA, a medium effect (Cohen's $f = 0.25$) with the same α and power requirement also needed a minimum $N \approx 52$. Accordingly, the ultimate sample ($N = 57$) breached both cutoffs, indicating adequate statistical power. The final participants were 57 intermediate EFL learners who attended a TOEFL preparation class at Neek Segal Language Institute in Tehran, Iran. They were enrolled through course announcements, and participation was voluntary, with written informed consent. Two intact classes were assigned to the experimental ($n = 28$; 11 male, 17 female) and control ($n = 29$; 10 male, 19 female) groups via convenience sampling technique. They were aged 22-35 years (experimental group: $M = 27.5$, $SD = 3.8$; control group: $M = 28.2$, $SD = 4.1$).

All participants were monolingual Persian speakers with intermediate English proficiency, as measured by the Oxford Placement Test (OPT; Allan, 1992), to ensure group homogeneity in terms of English proficiency. For the OPT intermediate level, the band score ranges from 135 to 150. Pretreatment equivalence based on age, OPT, reading test, and SORS test was examined prior to the intervention. There was no attrition, and data integrity was 100% over all phases. The intervention was 10 weeks in length and included 20 sessions (twice a week, 60 mins each), consistent with the institute's curriculum. Table 1 displays the demographic characteristics of the participants. To prevent the diffusion of treatment, classes met on alternate days in different rooms, and students were requested not to share materials across classes.

Table 1
Demographic Characteristics of Participants

Group	n	Age (M \pm SD)	Gender	Native Language	English Proficiency Level (Test)
Experimental	28	27.5 \pm 3.8	11 Male 17 Female	Persian	Intermediate (OPT)
Control	29	28.2 \pm 4.1	10 Male 19 Female	Persian	Intermediate (OPT)
Total	57	22–35 years	21 Male 36 Female	Persian	Intermediate (OPT)

3.3. Materials and Instruments

3.3.1. Materials

Eighteen original academic reading extracts (350-450 words each), were used. The passages were taken from the TOEFL iBT examination (Educational Testing Service, 2021). Readability, measured using the Flesch-Kincaid formula, was satisfactory for intermediate learners. Participants received (a) the passage, (b) five comprehension items, and (c) the LOA rubric during session packets for self- and peer-assessment. The LOA rubric categorized performance into three levels ("Correct," "Partially Correct," "Incorrect") with explicit strategy evidence indicators such as "cites relevant sentence for scanning; main idea vs. detail." The clarity of the rubric was piloted with five students and revised for wording.

3.3.2. Instruments

The first instrument was the Oxford Placement Test (OPT), comprising the listening and grammar items with multiple-choice format. The listening section has 100 items requiring learners to extract information from short audio clips, and the grammar includes 100 cloze items with a three-option multiple-choice format. According to the official OPT manual, scores between 135 and 150 are intermediate range scores. A preliminary reliability was carried out in a small sample of 30 Iranian learners (non-participants) and confirmed by the internal coherence ($KR-20 = .86$). In this study, OPT was only used as an initial screening and clustering variable, and the scores were not used as a dependent variable. Testing was conducted in a controlled lab setting with a 50-minute time constraint.

The second instrument was the Survey of Reading Strategies (SORS; Mokhtari & Sheorey, 2002), a 30-item self-report scale assessing English reading strategy use. The SORS has three scales of 13 items for global reading strategies, 8 items for problem-solving strategies, and 9 items for support strategies. Responses were made on a 5-point Likert scale (1-never to 5-always), resulting in a final score of 30-150. The internal consistency of the SORS has been found to be strong ($\alpha = .82-.89$) (Sheorey & Mokhtari, 2001). To reduce self-report bias, participants responded anonymously in coded envelopes and were asked to report actual, not ideal, behaviors. Despite strong empirical support for this measure, its reliance on a self-report format is a weakness, in that it does not capture online processing. This limitation was compensated by the triangulation of survey results with qualitative interview data.

The third instrument was a researcher-made reading comprehension test that was given as a pretest and posttest. Each test included four passages (380-450 words) selected from Bowling (2024). They were accompanied by 20 multiple-choice items. Two parallel versions were created to minimize practice effects, with passages matched for length, readability (Flesch-Kincaid), and topic familiarity. Content validity was confirmed through two practicing ELT faculty experts by assessing whether the strategies were relevant for the intermediate level. A pilot study with 15 students (not in the main sample) informed item analysis. Although not based on a large pilot scale, an item analysis was conducted on difficulty and discrimination indices. Weak features were removed or replaced. Reliability levels were adequate

(KR-21 = .82; parallel-form reliability $r = .83$). Item keys were double-checked by two additional reviewers. The duration of each test was 30 minutes.

Finally, to explore learners' perceptions of LOA, a subset of 10 learners from the experimental group was interviewed via semi-structured interviews. While the basic protocol included some closed questions (e.g., "Did the activity help?"), the items were modified to be open-ended (e.g., "Considering your reading skills, how were the activities affected?"). The last version of the protocol was also checked for clarity of expression and cultural sensitivity by two ELT teachers and trainers with more than 15 years of experience. The interviews were conducted in Persian; each lasted approximately 10-15 minutes and was conducted one week after posttests. The Persian responses were translated to English through a back-translation procedure (Brislin, 1986) by two bilingual research assistants who held PhDs in ELT to preserve meaning. Thematic analysis was conducted using Braun and Clarke's (2006) reflexive approach. No novel themes were observed at participant 10, indicating saturation was achieved. For reliability cross-checks, two trained coders analyzed the entire data separately, inter-coder agreement was calculated (Cohen's $\kappa = .89$), and disagreements were resolved through discussion. Coding and data management were supported by NVivo software.

- 1) In what ways did self-assessment assist you in your reading performance?
- 2) What did you like about giving and receiving feedback from your peers?
- 3) In what ways did the class exercises encourage you to become more conscious about your reading?
- 4) How was this class different from your past experiences with reading tests?
- 5) What are some reading strategies that you use now that you did not use before?

3.4. Procedure

First, the participants were chosen on a convenience basis from two intermediate EFL classes at the Neek Segal Language Institute in Tehran and were assigned to the experimental ($n = 28$) and control ($n = 29$) groups. During the pretest phase, all participants were given a pretest battery to measure baseline level of skill and to compare the two groups. These were the OPT for

intermediate-level proficiency, a reading comprehension test, and the SORS. Time-on-task was equated between the two groups (60 minutes/session).

The intervention group received an LOA program for a 10-week period. Each session consisted of a 60-min reading passage (e.g., renewable energy, cultural globalization) of 350-450 words and a five-item multiple-choice comprehension quiz. The teacher demonstrated the six reading strategies explicitly: predicting, scanning, skimming, inferencing, summarizing, and understanding the writer's tone (e.g., "Let's skim the first paragraph to find out what the main idea is"). Learners self-rated their answers on the three-option rubric following the quiz. For each item, participants recorded a justification for their self-rating in a notebook ("I scanned for dates to answer question 5") or an uncertainty ("I couldn't infer the author's tone in question 2"). This phase lasted for 20 minutes, during which time the teacher roamed the room to explain to students what needed to be completed for the rest of the activity and validate strategies used (e.g., "Loved your context clue strategy here!"). After self-rating, the students engaged in anonymous peer-rating. Answer sheets were randomly distributed, and the classmates graded answer sheets with the same rubric. Teacher-peer rating disagreements were discussed in teacher-facilitated conversations when the teacher presented de-identified exemplars of student work and demonstrated use of the rubric (for example, "This peer missed paragraph three-detail support, so it's 'Incorrect'").

Weekly collaborative tasks consisted of small-group discussions (e.g., "How did summarizing assist you in responding to question 7?"). Teacher-learner interactions focused on strategy. The teacher-researcher gave strategy-based feedback after checking the students' notes to monitor their progress (e.g., "You read more accurately when you skimmed this time!"). Peer- and self-assessment papers (strategy logs, rubric sheets) were gathered and scanned to enable a piece of process analysis; however, these were not graded and did not factor into course grades. The session format was fixed: 10 minutes deploying a strategy for study; 15 minutes reading the assigned text from which they were preparing to learn; 10 minutes for a quiz; 20 minutes for self and peer evaluation; 5 minutes for whole-class feedback.

During the intervention, teacher-learner interaction was lively, formative, and strategy-based in the experimental class. Instead of teaching from above, the teacher took a facilitating position and interacted with students when they were self- and peer-assessing. With students writing rationales for

their self-assessments, the instructor roved and individually coached, asking written-product-specific questions, reiterating the criteria for success on each rubric, and modeling effective reading strategy use (e.g., “Your ability to use context clues here was impressive”). During peer-rating periods, the teacher attempted to resolve disagreements through whole-class discussion by sharing anonymous instances of student reports to demonstrate the application of the rubric and to elicit critical self-reflection. Feedback was ongoing and targeted at students’ thinking rather than just at their performance. These interactions promoted a social classroom environment where assessment was used not only as a tool of measurement but also as a way of scaffolding learning and developing metacognitive awareness.

For the control group, traditional strategy instruction targeting the same six reading strategies (i.e., predicting, scanning, skimming, inferencing, summarizing, and understanding the author’s tone) was taught for the 10-week-long period, but the instruction was teacher-centered. Each 60-min lesson contained a 350-450-word passage and a 10-item quiz that matched the content for the treatment group. Using direct explanation (e.g., “Skimming means reading quickly for the main idea”) and modeling (e.g., “Here’s how I used context clues to guess ‘sustainable’”), the teacher-researcher (instructor) specifically taught the six strategies.

Learners were given summative feedback (e.g., “You got 7/10”) without self/peer-rating tasks after quizzes. Instruction focused on teacher-led practice: the teacher provided a review of answers aloud (e.g., “Question 4 asks you to look for numbers in paragraph 2”), and learners copied model responses. Interaction between teacher and learners was mostly one-way: transferring knowledge and dealing with errors and little reflection or discussion. The instructor went over right answers out loud, provided rationales for items, and requested that students copy model answers. Activities centered on individual strategy drills, such as timed practices (e.g., “Read this passage in 2 minutes and write the main idea”) or vocabulary packets (e.g., “Use context to define the bolded terms”). Critically, the total instructional time per session was equated with that of the LOA group; the duration allocated to self/peer assessment in the experimental condition was filled with more guided practice in the control condition to maintain a time match. In the control condition, no rubric use or peer-to-peer evaluation took place.

Posttest assessment was conducted one week after the intervention. The reading comprehension test (parallel passages) and SORS were administered again to both groups to assess the changes in strategy use. In the experimental group, semi-structured interviews (10 participants, 10-15 min each) were also conducted to investigate participants' perceptions of LOA. The posttest was applied under the same conditions as the pretest, and answer sheets were also anonymized and rated similarly.

3.5. Data Analysis

To answer the first and second research questions, paired-samples t-tests tested within-subject differences, whereas independent-samples t-tests were used to compare post-intervention results across groups. Moreover, group \times time interactions, involving between-subjects and within-subjects factors, were analyzed across the three SORS subscales simultaneously using mixed design MANOVA.

To investigate learners' attitudes toward LOA, semi-structured interviews were held with 10 individuals in the experimental group. Even though the original protocol included a number of closed questions (e.g., "Did the activities help?"), items were reworded for open-ended responses (e.g., "How did the activities affect your reading?"). The final protocol was checked by two ELT teachers who had worked for more than 15 years to improve its clarity and cultural adaptability. Analysis of data was conducted using Braun and Clarke's (2006) framework. The point of data saturation was determined to be when no new themes emerged; the result was at the 10th participant. To increase reliability, a full clean dataset was independently coded by two trained coders, and inter-coder agreement was used. Discrepancies were resolved by consensus. Coding and data management were supported by NVivo software.

4. Results

1.1. Results of Quantitative Phase

Parametric assumptions were tested prior to hypothesis testing. Normality was assessed by the Shapiro-Wilk test. All variables were approximately normally distributed at pretest and posttest (reading comprehension: experimental, $p = .28$; control, $p = .31$; SORS Global: experimental, $p = .42$; control, $p = .33$; SORS problem-solving: experimental, $p = .48$, control, $p = .29$; SORS support: experimental $p = .35$, control $p = .41$). Homogeneity of variances were assessed using Levene's test, which yielded

nonsignificant results across all between-group comparisons (all p values > .05). In the case of SORS, equality of covariance matrices was examined with Box's M, which was nonsignificant ($p = .16$). Table 2 shows the descriptive statistics (means, SDs, minimum, and maximum) for reading comprehension and SORS scores at pretest and posttest.

Table 2
Descriptive Statistics for All Variables by Group and Time

Variable	Group	Time	M	SD	Min	Max
Reading Comprehension	Experimental	Pre	12.45	2.31	8	18
Reading Comprehension	Experimental	Post	16.82	1.97	12	20
Reading Comprehension	Control	Pre	12.39	2.28	7	17
Reading Comprehension	Control	Post	13.91	2.14	9	18
SORS Global Strategies	Experimental	Pre	32.10	5.45	20	45
SORS Global Strategies	Experimental	Post	45.25	4.82	35	60
SORS Global Strategies	Control	Pre	31.85	5.67	19	44
SORS Global Strategies	Control	Post	34.12	5.03	22	47
SORS Problem-Solving	Experimental	Pre	19.80	3.51	12	28
SORS Problem-Solving	Experimental	Post	31.65	3.22	25	38
SORS Problem-Solving	Control	Pre	19.45	3.49	11	27
SORS Problem-Solving	Control	Post	21.04	3.58	14	30
SORS Support Strategies	Experimental	Pre	22.50	4.56	14	35
SORS Support Strategies	Experimental	Post	36.20	4.10	28	42
SORS Support Strategies	Control	Pre	22.15	4.75	13	34
SORS Support Strategies	Control	Post	24.89	4.63	16	37
SORS Total	Experimental	Pre	74.40	10.22	57	105
SORS Total	Experimental	Post	113.10	9.75	98	134
SORS Total	Control	Pre	73.45	11.28	55	103
SORS Total	Control	Post	80.05	10.64	64	109

As Table 2 shows, no significant intergroup differences were found in pretest reading comprehension results. However, the experimental group made significantly greater gains on the posttest measure of reading comprehension ($M = 16.82$, $SD = 1.97$) and SORS subscales (e.g., problem-solving: $M = 31.65$, $SD = 3.22$) compared to control group (reading comprehension $M = 13.91$, $SD = 2.14$).

To address the first research question, paired and independent samples t-tests were conducted on the reading comprehension pretest and posttest scores (Table 3). Equality of variances was checked using Levene's test ($p > .05$).

Table 3*Paired and Independent Samples t-Tests for Reading Comprehension*

Analysis Type	Group	<i>M</i> (Pre)	<i>M</i> (Post)	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Within-Group	Experimental	12.45	16.82	13.72	27	<.001	1.85
Within-Group	Control	12.39	13.91	3.01	28	.006	0.63
Between-Group (Pre)	Experimental vs. Control	12.45 vs. 12.39		0.09	55	.931	0.03
Between-Group (Post)	Experimental vs. Control	16.82 vs. 13.91		5.14	55	<.001	1.41

The experimental group showed a significant difference between the pretest and the posttest. Nonetheless, there were substantial between-group differences at posttest, with the participants in the intervention condition exhibiting significantly greater scores ($M = 16.82$) than the control group ($M = 13.91$), $t(55) = 5.14$, $p < .001$, $d = 1.41$.

To address the second research question, a mixed factorial MANOVA was conducted with time as the within-participant factor and group (experimental group vs. control group) as the between-participant factor for the three SORS subscales (i.e., global, problem-solving, and support strategies).

Table 4*Mixed-Design MANOVA Results for SORS Subscales*

Effect	Wilks' Λ	<i>F</i>	<i>df</i>	<i>p</i>	partial η^2
Time	.15	102.58	3, 53	<.001	.85
Group	.35	33.47	3, 53	<.001	.66
Time \times Group	.30	41.26	3, 53	<.001	.70

Table 4 displays the SORS results for reading strategy use. The mixed-design MANOVA showed a significant interaction of group and time, $F(3, 53) = 41.26$, $p < .001$, Wilks' $\Lambda = .30$, partial $\eta^2 = .70$. It means that the development of strategy use over time varied considerably between groups. Multivariate main effects of time ($F(3, 53) = 102.58$, $p < .001$, $P\eta^2 = .85$) and group ($F(3, 53) = 33.47$, $p < .001$, partial $\eta^2 = .66$) were also significant. Univariate follow-up tests were run following the significance of the multivariate interaction identified in the mixed-design MANOVA and are shown in Table 5. These analyses evaluate whether the group \times time interaction was significant for the three SORS subscales: global strategies, problem-solving strategies, and support strategies.

Table 5*Univariate Tests of Group \times Time Interactions for SORS Subscales*

Subscale	<i>F</i>	<i>df</i>	<i>p</i>	partial η^2	Posttest <i>d</i>
Global Strategies	65.49	1, 55	<.001	.54	1.12
Problem-Solving	97.61	1, 55	<.001	.64	1.35
Support Strategies	89.26	1, 55	<.001	.62	1.29

As shown in Table 5, the interaction effect for global strategies was statistically significant, $F(1, 55) = 65.49$, $p < .001$, $P\eta^2 = .54$, indicating a large effect. The increase in global strategy use across time was substantially greater from pre-intervention to post-intervention for the experimental group than for the control group. This advantage was further supported by a between-group posttest comparison, with a large effect size (Cohen's $d = 1.12$).

Additionally, for problem-solving strategies, the group \times time interaction was also significant, $F(1, 55) = 97.61$, $p < .001$, $P\eta^2 = .64$. This indicates that the amount of problem-solving strategies used by LOA participants increased substantially relative to the control group. The size of the difference at posttest was also very large (Cohen's $d = 1.35$), meaning that LOA significantly assisted learners to apply active strategies (e.g., inference-making, contextual guessing, and self-monitoring). The corresponding interaction was also significant and large, $F(1, 55) = 89.26$, $p < .001$, partial $\eta^2 = .62$. Moreover, experimental group learners significantly increased their use of support strategies, but the control group gained little. Similarly, the posttest effect size was large (Cohen's $d = 1.29$), indicating that LOA maintained its advantage in this area.

4.2. Results of Qualitative Phase

NVivo 12 was used to analyze the interview transcripts ($n = 10$). Interviews were coded according to the interview guide (deductive codes, e.g., "self-assessment," "peer feedback"), and a codebook was developed through iterative open coding (inductive codes, e.g., "assessment anxiety," "strategy experimentation"). Two trained coders applied the codebook to the entire dataset, yielding high inter-coder reliability (Cohen's $\kappa = .89$). Table 6 shows the theme, description, prevalence, and illustrative quotations.

Table 6
Themes and Representative Participant Comments

Theme	Prevalence	Description	Representative quotations
Metacognitive awareness	7/10	Reflection on comprehension and error patterns	“When I checked why my answer was wrong, I realized I skipped a clue word.” (P1)
Strategic reading development	8/10	Adoption/refinement of strategies (e.g., skimming, inferencing)	“I started skimming before reading the questions, which changed my answers.” (P3)
Peer collaboration and feedback	10/10	Learning from peers’ strategies and reassurance	“Seeing my classmate’s notes showed me another way to find the answer.” (P2)
Transformative view of assessment	8/10	Shift from punitive to formative framing	“Before, tests were frightening. Here, they became part of learning.” (P5)
Motivation and engagement	7/10	Increased investment due to active involvement	“I felt responsible when I had to justify my answers—it kept me focused.” (P6)

As Table 6 illustrates, the first theme, metacognitive awareness, represents the degree to which students became more reflective about their reading processing. The results also showed that seven participants felt able to monitor and evaluate their comprehension through the use of a rubric guide and justification of their answers. The deeper internalization of the strategic thinking process is reflected in one student’s remark: “When I looked at my mistakes, I realized where I wasn’t thinking right.” This reflection allowed learners to discern particular limitations in their reading, for example, misinterpreting inferential questions or ignoring supportive cues.

The second theme, strategic reader development, indicates that students not only developed a greater sense of awareness for using strategies but also started utilizing new strategies. Six participants referred to the change in which they dealt with the texts, such as the order in which they read them or the kind of attention they paid to question stems. One participant stated, “Now I skim the text first before reading the questions, which I didn’t do before,” reflecting enhanced pre-reading behavior. The flexibility of LOA seemed to develop a learning cycle that was able to enhance their strategic ability.

The third theme, peer collaboration and feedback, sheds light on the social aspect of the learning that was triggered by the invitation for peer rating and discussion. All respondents valued the opportunity to both assess and be

assessed by their peers. Discussing with a peer also revealed alternative strategies, as one student explained, "My friend used scanning to obtain the answer quickly, so I tried that." These exchanges rendered the learning experience more dialogic and less didactic.

The fourth theme is transformative perceptions of assessment, and this concept represents a major shifting of cognitive and emotional gear for participants in the project. Eight participants stated that unlike their previous high-stakes testing experiences, LOA lessened their anxiety and presented assessments as a motivator for growth. As one interviewee put it, "I used to have a fear of tests before this course. And now, I think about them as a part of learning." They broke the cycle by not giving learners some scores but granted them immediate, explainable feedback. This change in focus may have also helped foster a more positive and less stressful learning atmosphere.

Finally, the motivation and engagement theme emerged with learners' continued participation and active engagement over the 10-week intervention. It made the lessons more interactive and meaningful. One student remarked, "It was more interesting than ordinary classes because we did something different," and another reported, "I could feel more involved in class because I played a role in verifying my answer." These statements indicate cognitive engagement and emotional involvement were promoted by the LOA model.

Taken together, the qualitative results indicate that LOA changed the ways that learners engaged with reading comprehension and strategy use. It promoted the awareness, cooperation, and skill of performing constructive assessment. However, beyond their correspondence to the extent of the quantitative gains in the development of reading performance and strategy use, these results provide evidence of how pedagogical intervention in the design of assessment can affect learners' identities, agency, and motivation.

5. Discussion

The results suggest that the integration of LOA in EFL reading instruction can contribute to substantial gains in reading comprehension and strategy use. Comprehension gains and significant increases in metacognitive awareness of reading strategies were observed among participants who received the LOA intervention compared to their control group counterparts. Although this is consistent with previous work (e.g., Beikmohammadi et al., 2020; Yan & Carless, 2022), it is of particular significance due to the emphasis on formative, interactive assessment processes. Learners were not simply

assessed to read. They were asked to think, self-evaluate, and collaborate with friends, so learning was a collective and ongoing effort. Learning is best mediated through social and scaffolded interaction (Vygotsky, 1978). Here, self-assessment and peer assessment work as mediating means, as students need to internalize the knowledge. These improvements suggest that exposure to standard reading instruction and to repeated testing alone may drive incremental learning effects, which were also observed in Banitalebi and Ghiasvand's (2023) study. Therefore, LOA does not appear to be the only contributing factor to growth in comprehension; rather, it seems to catalyze and intensify gains that assistance is also likely to provide. A further reason could be attributed to the Hawthorne effect, such that experimental group participants, who knew that they were being treated specially, made more of an effort. There may have been teacher bias, as the same teacher delivered the LOA intervention and traditional testing. This potential confound underscores the importance of replication using blind raters and independent instructors.

It also emerged that learners started to reconsider how they read. They reported being aware of holes in their understanding, for example, and trying out strategies such as scanning and inferring and monitoring their understanding. Such observations are highly consistent with metacognitive control (Turner & Purpura, 2015). With LOA, while the strategy instruction itself was presented, the learner was also prompted to practice and to reflect on the strategies in discussion and in feedback. This view bolsters Grabe's (2009) argument that successful reading depends not only on having strategies but also on the appropriate use of them in real situations.

Notably, due to their developmental level, EFL students generally do not have the required metacognitive awareness for the successful implementation of such reading strategies (Ghaneiarani et al., 2024), and our findings demonstrated that structured LOA practice can help them develop that awareness. Self-rating activities encouraged learners to thematize their thinking and make explicit patterns in their reading behavior, which is why they developed in the direction of more self-monitoring. As one participant articulated, "I tend to read more carefully now, especially since I saw one error in my self-rating, and now I feel more confident," a salient feature of LOA: that of turning mistakes into learning experiences. This self-correcting cycle renders LOA dissimilar to any summative assessment models, which seldom accommodate room for such reflection. Learners' reading development was also significantly affected by peer collaboration. The results of the study indicate that peer assessment was not a procedural technicality but a critical pedagogical event. Additionally, learners frequently mentioned that they re-

evaluated their own interpretation based on their exposure to a peer's interpretation.

The mixed-design MANOVA showed group \times time interactions for all three SORS subscales. Both the breadth and depth of strategy use increased by LOA learners as evidenced by positive gains in global, problem-solving, and support strategies among LOA group learners. These are reminiscent of the model proposed by Turner and Purpura (2015) based on metacognitive regulation, in which students develop greater metacognitive awareness and control over strategy use. However, the self-reported nature of the source of information raises concern about the accuracy because students may overestimate how much they use a strategy, especially if they are given explicit strategy instruction (Mokhtari & Sheorey, 2002). Interestingly, the control group demonstrated minimal transfer of strategy use, but there were no differences in the support strategies reported by the experimental and control groups. This implies that regardless of guided LOA, some learners may form strategies, such as using a dictionary or rereading, when tested multiple times.

The results replicate and extend the literature on LOA. Like Khalili et al. (2024), who documented teachers' greater awareness of student progress, the current study also points to commensurate benefits for learners. However, unlike research in which LOA brought about only metacognitive gains with no significant increase in test scores (Banitalebi & Ghiasvand, 2023), our study exhibits effects on both the process and product of learning. One explanation might be the amount of the intervention: LOA was implemented in almost every lesson, not haphazardly. It indicates that fidelity of implementation, the degree to which LOA pervades classroom instruction, may be the factor that determines whether gains are observable in performance tests.

6. Conclusions and Implications

This study aimed to explore LOA in Iranian EFL learners' reading comprehension and their use of reading strategies. Participants in the LOA group performed significantly better on comprehension tasks than those who experienced more traditional forms of assessment. However, the results must be taken with great caution. The improvements are not just due to LOA, as teacher enthusiasm, novelty of the approach, or an enthusiastic learner (the Hawthorne effect) could also have contributed. The sample size of this context-specific study was small, and the same instructor taught both groups, which may have caused bias. In addition, self-reported strategy use may lead to overestimated strategy use, as well as the 10-week intervention that did not

allow the researchers to assess whether the benefits would be maintained and transferred to other situations.

Nevertheless, the present study does provide information about how LOA principles might be used in practice. Instead of treating assessment as independent or an ending process, it takes into account assessment in the realm of instruction, including tasks such as self-rating with rubric, peer feedback on comprehension process, and post-task reporting. These activities promoted learners to self-monitor their strategy use and engage in error analysis, as well as work with peers, which is consistent with the more general objective of encouraging autonomy and agency in reading. Nonetheless, the extent to which there can be broader implementation hinges on institutional realities. Large class sizes, exam-focused syllabi, and low teacher preparation, issues observed in Iranian and many other EFL settings, may limit the extent to which the LOA model can be scaled without further support.

Future studies hence need to build on our current findings by investigating whether LOA-induced benefits persist for a longer duration than a transitory training experience, whether they would be transferred to other high-risk reading passages such as TOEFL and IELTS reading sections, and whether they would be replicated in classrooms in which peer assessment is under different cross-cultural orientations. Adding behavioral evidence in the form of eye-tracking or think-aloud data to self-report evidence would add credibility to claims regarding metamemory development. In addition, digital tools like AI-empowered feedback platforms offer potential avenues for scaling LOA practices in manners that reduce teacher workload while maintaining formative depth.

References

- Al-Abri, A., Ranjbaran Madiseh, F., & Morady Moghaddam, M. (2025). Exploring learning-oriented assessment in enhancing students' lexical fluency through MALL. *The Asia-Pacific Education Researcher*, 34(1), 1-13. <https://doi.org/10.1007/s40299-024-00832-7>
- Allan, D. (1992). *Oxford placement test*. Oxford University Press.
- Banitalebi, Z., & Ghiasvand, F. (2023). The representation of learning-oriented assessment (LOA) practice in EFL contexts: A case study of teacher questioning strategies. *Language Related Research*, 14(3), 69-95. <https://doi.org/10.29252/LRR.14.3.3>
- Beikmohammadi, M., Alavi, S. M., & Kaivanpanah, S. (2020). Learning-oriented assessment of reading: A mixed methods study of Iranian EFL university instructors' perceptions and practices. *Journal of Foreign Language Research*, 10(2), 316–329. <https://doi.org/10.22059/jflr.2019.264150.541>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Bowling, M. (2024). *TOEFL iBT 2024-2025 Preparation Book-TOEFL Secrets Study Guide*. Mometrix publication.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Sage.
- Brown, H. D. (2018). *Teaching by principles: An interactive approach to language pedagogy (4th ed.)*. Pearson Education.
- Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge. <https://doi.org/10.4324/9780203128213>
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963–976. <https://doi.org/10.1007/s10734-014-9816-z>
- Creswell, J. W., & Creswell, J. D. (2022). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (6th ed.). Sage.
- 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(1), 1–18. <https://doi.org/10.1186/s40468-022-00204-2>
- Educational Testing Service. (2021). *The official guide to the TOEFL iBT test* (6th ed.). McGraw-Hill Education.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Ghaneiarani, S., Alavi, S. M., & Kaivanpanah, S. (2024). Enhancing writing ability of Iranian EFL learners through learning-oriented assessment: Peer and teacher feedback implementation. *Language Testing in Asia*, 14(1), 39. <https://doi.org/10.1186/s40468-024-00298-w>
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Jalilzadeh, K., & Coombe, C. (2023). Constraints in employing learning-oriented assessment in EFL classrooms: Teachers' perceptions. *Language Testing in Asia*, 13(1), 7–25. <https://doi.org/10.1186/s40468-023-00222-8>
- Jones, N., & Saville, N. (2016). Learning oriented assessment: A systemic approach. *Studies in Language Testing*, 45(4), 23–47.
- Khalili, A., Zafarani, P., & Gholami, J. (2024). Learning-oriented assessment in the context of Iran: Teachers' perspectives. *International Journal of Language Testing*, 14(2), 82–96. <https://doi.org/10.22034/ijlt.2024.433024.1317>
- Mokhtari, K., & Sheorey, R. (2002). Measuring ESL students' awareness of reading strategies. *Journal of Development Education*, 25(3), 2–10.
- Nurjamin, A., Salazar-Espinoza, D. E., Saenko, N., & Bina, E. (2023). Learner-oriented assessment matters: Testing the effects of academic buoyancy, reflective thinking, and learner enjoyment in self-assessment and test-taking anxiety management of the EFL learners. *Language Testing in Asia*, 13(1), 1–30. <https://doi.org/10.1186/s40468-023-00247-z>
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(1), 190–208. <https://doi.org/10.1111/modl.12308>
- Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, 29(4), 431–449. [https://doi.org/10.1016/S0346-251X\(01\)00039-2](https://doi.org/10.1016/S0346-251X(01)00039-2)
- Turner, C. E., & Purpura, J. E. (2015). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari, & J. Banerjee (Eds.),

- Handbook of second language assessment* (pp. 255–273). De Gruyter Mouton.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Massachusetts: Harvard University Press.
- Yan, Y., & Carless, D. (2022). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, 47(7), 1116–1128.
<https://doi.org/10.1080/02602938.2021.2001431>
- Zhang, L. J. (2001). Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an acquisition-poor environment. *Language Awareness*, 10(4), 268–288.
<https://doi.org/10.1080/09658410108667039>