

# The Effect of Learning-Oriented Assessment on EFL Learners' Peer and Self-Rating Accuracy of Reading Comprehension Tasks: A Mixed-Methods Study

Pegah Doroudi<sup>1</sup>, Mohammad Taghi Farvardin<sup>\*2</sup>, Ghafour Rezaie Golandouz<sup>3</sup>

<sup>1</sup>Ph.D. Candidate, Department of Foreign Languages Teaching, SR. C., Islamic Azad University, Tehran, Iran

<sup>2</sup>Assistant Professor, Department of Foreign Languages Teaching, Ahv. C., Islamic Azad University, Ahvaz, Iran

<sup>3</sup>Assistant Professor, Department of Foreign Languages Teaching, SR. C., Islamic Azad University, Tehran, Iran

DOI: [10.71864/LCT-2025-1214257](https://doi.org/10.71864/LCT-2025-1214257)

Received: 02/06/2025

Revised: 03/09/2025

Accepted: 06/09/2025

Published: 07/09/2025

**Keywords:** Learning-oriented assessment; Peer rating accuracy; Perceptions of assessment; Reading comprehension; Self-rating accuracy

**Abstract:** Learning-oriented assessment (LOA) is a method that integrates assessment into the educational process, with a focus on formative feedback and collaborative engagement. However, the practical efficacy of LOA in improving the accuracy of self- and peer ratings in an English as a foreign language (EFL) context has not yet been examined. Therefore, this study, following an explanatory sequential mixed-methods design, explored the role of LOA in the self- and peer rating accuracy of intermediate EFL learners. To this end, 57 intermediate EFL learners were non-randomly divided into an experimental and a control group. Over a 10-week intervention, participants in the experimental group (n = 28) completed LOA tasks, including rubric-guided self-/peer-rating of reading comprehension and strategy-oriented feedback; the control group (n = 29) followed conventional summative assessments for the same period. Pre/posttests self- and peer rating accuracy measures ( $\Delta$  scores = |learner rating – teacher rating|) were calculated using independent and paired samples t-tests. Ten semi-structured interviews were also thematically analyzed. The interviews were coded by two coders, and then the inter-coder agreement was calculated. The Cohen's kappa coefficient of .89 indicated acceptable reliability. The quantitative analysis showed that the experimental group's self-rating and peer rating accuracy significantly improved. The control group, however, showed no significant change.

\* Corresponding Author's E-mail address: [mtfarvardin@iau.ac.ir](mailto:mtfarvardin@iau.ac.ir)



This work is licensed under a [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/).

**Cite this article:** Doroudi, P., Farvardin, M. T., & Rezaie Golandouz, Gh. (2025). The Effect of Learning-Oriented Assessment on EFL Learners' Peer and Self-Rating Accuracy of Reading Comprehension Tasks: A Mixed-Methods Study. *Journal of Language, Culture, and Translation*, 8(1), 102-125. <https://doi.org/10.71864/lct-2025-1214257>

Additionally, the thematic analysis revealed five emergent themes: self-rating as a tool for metacognitive growth, peer rating as a collaborative learning opportunity, ambiguity in rubric application, emotional barriers to honest self-assessment, and desire for structured guidance. The findings indicate that LOA can significantly enhance rating accuracy.

## 1. Introduction

Grounded in sociocultural theory (Vygotsky, 1978), learning-oriented assessment (LOA) regards assessment as a mediator of learning, scaffolding skills through social interaction and iterative practice, principles especially relevant to the complex demands of EFL reading comprehension (Carless, 2012). In conventional assessments, standardized testing is usually given top priority; more often than not, it evaluates outcomes rather than encourages critical reading of texts (Al-Abri et al., 2025). Within the LOA framework, reading comprehension is seen as an evolving construct supported through ongoing formative interactions rather than as a fixed endpoint. Students who actively assess their knowledge, by self- and peer assessments, begin to absorb the skills and techniques that support more complex knowledge (Tsagari & Banerjee, 2016). LOA can help students track their reading processes, revise interpretations, and apply more successful strategies using dialogic feedback and scaffolded practice, resulting in a better understanding (Yan & Carless, 2022). Through teacher-learner interactions, dynamic assessment (DA), a subset of LOA, combines instruction and assessment to give students scaffolded support to internalize reading strategies (Al-Abri et al., 2025; Ghaneiarani et al., 2024; Khalili et al., 2024; Khodami, 2023; Zeinali et al., 2025).

Self-assessment, a pillar of LOA, enables students to examine their growth critically, thus fostering natural motivation and autonomy (Al-Abri et al., 2025; Nazari, 2023). Particularly when matched with proficiency levels, studies in Iranian EFL environments show that structured self-assessment tools increase learners' metacognitive awareness and accuracy in reading tasks (Khalili et al., 2024). However, the success of self-assessment depends on students' ability to internalize criteria and apply them reflectively, a process that calls for both explicit instruction and teacher direction (Beikmohammadi et al., 2020). Peer assessment is another pillar of LOA since it helps students to exchange useful comments, clear ambiguity, and model successful strategies, promoting cooperative learning (Liu & Carless, 2006).

A key component of LOA is its conversion from conceptual knowledge into practical classroom implementation (Jones & Saville, 2016). While

theories of assessment literacy inform teachers of *what* assessment for learning involves, the actualization of LOA depends on *how* these principles are operationalized in practice (Hume & Coll, 2009). Teachers may face challenges in applying LOA, which can be because of the summative assessment dominance and unfamiliarity with well-designed LOA frameworks (Khalili et al., 2024). One of the practical frameworks for implementing LOA in the classroom has been proposed by Jones and Saville's (2016), which provides a practical roadmap for operationalizing assessment principles in instructional contexts. This model views the classroom as the focal point of a dynamic cycle of interaction between teaching, learning, and assessment. Students participate in peer and self-assessment supported by clear criteria and teacher scaffolding, while teachers create assignments that not only encourage the use of strategies and reflection but also elicit evidence of learning.

In LOA research, the accuracy of peer and self-rating has not been well-addressed (Al-Rashidi et al., 2022). Furthermore, little is known about how students view LOA when it is applied in classroom. In order to fill these gaps, the current mixed-methods study recorded Iranian EFL learners' opinions of LOA as a teaching strategy and investigated how LOA affected their rating accuracy on reading comprehension tasks. It specifically aimed to respond to the following research questions:

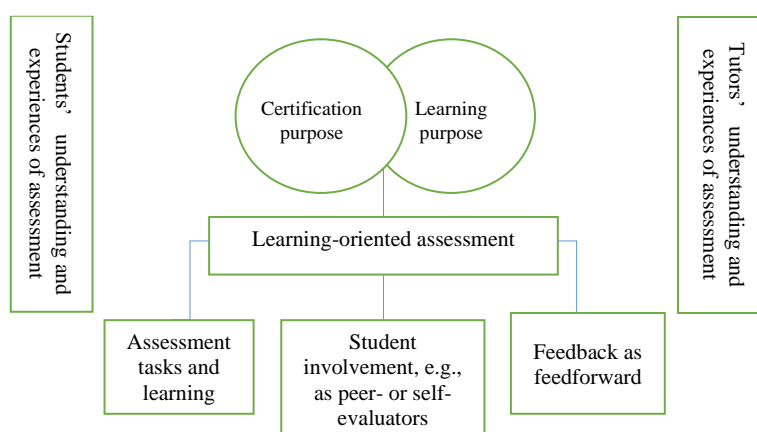
1. Does learning-oriented assessment significantly affect EFL learners' self-rating accuracy of reading comprehension tasks?
2. Does learning-oriented assessment significantly affect EFL learners' peer rating accuracy of reading comprehension tasks?
3. What are EFL learners' perceptions of learning-oriented assessment?

## 2. Literature Review

The idea of LOA first appeared in language testing research with Purpura (2004), who argued that assessment should be more than a tool for measurement; it should also support learning. Building on this, Carless (2007) gave LOA its more familiar definition as an approach that puts learning at the center. Later, Carless (2015) highlighted the processes of LOA, stressing that assessment must work as a mechanism for reflection and long-term learning, not simply as a judgment of outcomes. Jones and Saville (2016) extended this work by framing LOA as a systemic model in which assessment is woven into the context of classroom. It was an important shift from viewing LOA as a useful idea to treating it as a framework with practical, actionable steps for teachers and learners.

In second and foreign language classrooms, assessment has the potential to become a bridging device between teaching and learning (Carless, 2007). Starting from this premise, the LOA approach focuses on the learning process and aims at finding new ways for teachers and learners to benefit from the information generated by planned and unplanned assessment opportunities (Tsagari & Banerjee, 2016). Carless (2007) underlined the fact that in LOA, learning comes first, both in the way the term is literally constructed and as a matter of the principle of emphasizing the learning aspects of assessment, and introduces a framework for LOA (Figure 1).

**Figure 1.** Framework for Learning-Oriented Assessment (Carless, 2007)



In this model, the two main purposes of assessment are overlapping, meaning that efficient assessment is taking place; furthermore, the model identifies three interactive elements of LOA: (a) assessment tasks as learning tasks, (b) student involvement, and (c) feedback as feed-forward (Carless, 2007). Three principles summarize this model: Assessment tasks should be designed to stimulate sound learning practices amongst students; Assessment should involve students in actively engaging with criteria quality (their own and/or peers' performance); Feedback should be timely and forward-looking so as to support current and future student learning (Carless, 2007).

Among the well-known practical models of LOA is Jones and Saville's (2016) *classroom within an LOA model*. This model places the classroom at the center of the learning-assessment cycle, ensuring that every task serves both instructional and evaluative purposes. It offers five interconnected principles. First, assessment tasks must be aligned with curriculum goals, ensuring they reflect real learning outcomes. Second,

tasks should function simultaneously as opportunities for learning and as ways of collecting evidence of progress. Third, learners should engage actively with explicit criteria, developing assessment literacy by applying rubrics and benchmarks to their own and peers' work. Fourth, feedback should create ongoing loops that allow teachers and students to adapt instruction and strategies in real time. Finally, the model pays attention to the emotional side of assessment, structuring tasks and feedback in ways that build confidence and sustain motivation. These principles create a dynamic cycle, i.e., task design, learner engagement, feedback, adaptation, and refinement, through which assessment genuinely drives learning. In this way, Jones and Saville's (2016) model moves LOA from theory to practice, offering a roadmap for classrooms that is both structured and adaptable. Within this framework, self- and peer assessment have received particular attention. Research suggests that self-assessment, when supported with training and clear criteria, enhances metacognitive awareness and encourages learners to take ownership of their learning (Yan & Carless, 2022).

Empirical work in EFL settings offers a mixed but increasingly encouraging picture of LOA. For example, Al-Rashidi et al. (2022) found that LOA improved the participants' writing accuracy by encouraging collaboration, though it had limited effects on reading comprehension due to the absence of structured, reading-specific peer tasks. More recent studies have highlighted LOA's potential in other areas: Al-Abri et al. (2025) showed gains in lexical fluency through LOA-based mobile-assisted learning, while Ghaneiarani et al. (2024) found that teacher- and peer feedback cycles enhanced writing performance. In addition, Khalili et al. (2024) explored Iranian teachers' perspectives on LOA and found broad support for its benefits, though teachers noted significant implementation challenges, such as a lack of clear models, insufficient training, and limited resources.

Despite these advances, a clear gap remains. Most studies have focused on writing or oral skills, while the use of LOA to support reading comprehension and, in particular, the accuracy of self- and peer ratings in this domain has received little attention. Moreover, even when reading is studied, assessment practices are often fragmented rather than embedded in a systematic LOA cycle. Importantly, although Jones and Saville's (2016) classroom model provides a practical roadmap that integrates alignment, feedback, learner engagement, and affective support, few studies have used it as the foundation for research design. This means we still know little about how the model works when fully applied in classroom practice, especially in contexts such as Iran where traditional,

summative testing dominates. To address this methodological and contextual gap, the present study applies Jones and Saville's classroom model to examine how LOA influences Iranian EFL learners' self- and peer-rating accuracy in reading comprehension tasks, and how learners perceive this approach in practice.

### **3. Method**

#### **3.1. Research Design**

The three research questions were addressed using a sequential explanatory mixed-methods design that combined a qualitative and a quasi-experimental quantitative phase.

#### **3.2. Participants**

In this study, 57 intermediate EFL learners enrolled in a TOEFL preparation course at a language institute in Tehran, Iran, were recruited. Participants, aged 22–35 years (experimental group:  $M = 27.5$ ,  $SD = 3.8$ ; control group:  $M = 28.2$ ,  $SD = 4.1$ ), were selected through convenience nonrandom sampling, with two intact classes assigned to experimental ( $n = 28$ ; 11 male, 17 female) and control ( $n = 29$ ; 10 male, 19 female) groups. All participants were native Persian speakers and demonstrated intermediate English proficiency, confirmed via the Oxford Placement Test (OPT), which ensured homogeneity in English proficiency. The 10-week intervention comprised 20 sessions (two 60-minute sessions weekly), aligning with the institute's curriculum. The participants' demographic information is presented in Table 1. Before the study, informed consent was obtained from all participants through a detailed consent form outlining the study's purpose, confidentiality protocols, and voluntary participation terms.

#### **3.3. Instruments**

The first instrument used in this study was the OPT, which consists of listening and grammar sections in multiple-choice format (Allan, 1992). The listening section consists of 100 items. Test-takers are asked to choose the correct word, which they hear in short sentences, from two choices. The grammar section consists of 100 items. Test-takers are asked to read the stem with a blank and to choose one of the three options for the blank. The band score of 135 to 150 is considered the intermediate level of English proficiency. The OPT's test-retest reliability has been validated through repeated administrations, with learners' scores showing strong correlation ( $r = 0.82$ – $0.87$ ) over short intervals, confirming its temporal stability (Allan, 1992). These features ensure that the OPT

reliably distinguishes between proficiency levels, critical for accurate participant grouping in research contexts. Furthermore, criterion-related validity is demonstrated through strong correlations ( $r = 0.78\text{--}0.84$ ) between OPT scores and performance on other standardized tests (e.g., TOEFL) (Allan, 1992).

For the self-rating accuracy, after learners completed a 20-item reading comprehension pretest, they rated their own answers using a three-choice rubric (Correct = 2, Incorrect = 0, Not Sure = 1). The self-rating accuracy  $\Delta$  was calculated as the absolute difference between self-ratings and teacher scores (e.g.,  $|\text{self-score} - \text{teacher score}|$ ). For the peer rating accuracy task, participants anonymously rated a classmate's quiz using the same rubric, with the peer rating accuracy  $\Delta$  calculated similarly (See Appendix).

To investigate learners' perceptions of LOA, semi-structured interviews were conducted in both English and Persian with 10 participants from the experimental group. Ten participants were selected because the collected data showed data saturation after having interviews with the tenth participant. In addition, two PhD holders in English language teaching with more than 10 years of experience reviewed the interview questions to be relevant and comprehensible for the intermediate EFL learners. The interview lasted about 10 minutes for each participant. It included five open-ended questions as follows:

1. *How did you feel when you checked your own answers?*
2. *Was it helpful to check and score your classmates' answers? Why?*
3. *What were the problems with scoring and checking your own or your classmate's answers?*
4. *Did the scoring guide (like the 0-2 scale) help you understand your mistakes?*
5. *How can we improve these scoring activities?*

Questions were rephrased for clarity as needed (e.g., "*scoring guides*" instead of "*rubrics*"), and participants were encouraged to give examples. Interviews were audio-recorded, transcribed, and analyzed thematically. Data saturation was achieved with 10 participants, as no new themes emerged. While language barriers occasionally arose (e.g., hesitations or vocabulary gaps), simplified phrasing and patience ensured meaningful dialogue.

### 3.4. Materials

This study utilized 18 original academic reading passages, each ranging from 350 to 450 words, designed for intermediate EFL learners.



These passages were adapted from TOEFL iBT materials (Educational Testing Service, 2012) to ensure alignment with standardized academic texts while avoiding direct copyright infringement. Topics spanned four disciplines: social sciences (e.g., urbanization, migration), natural sciences (e.g., renewable energy, genetic engineering), humanities (e.g., cultural relativism, globalization), and technology. Each passage was calibrated for readability using Flesch-Kincaid Grade Level scores (8.15-9.36). As calculated by Flesch-Kincaid formula, the readability level of the reading passages was as follows: Passage one: 8.35, passage two: 8.47, passage three: 9.15, passage four: 8.75, passage five: 8.61, passage six: 8.92, passage seven: 9.33, passage eight: 9.05, passage nine: 8.84, passage ten: 8.57, passage eleven: 9.14, passage twelve: 8.60, passage thirteen: 8.29, passage fourteen: 8.53, passage fifteen: 9.27, passage sixteen: 9.36, passage seventeen: 8.15, passage eighteen: 8.42, which were considered satisfactory for intermediate students.

To ensure that assessment practices were pedagogically meaningful, the instructional design of the experimental group was grounded in Jones and Saville's (2016) classroom within an LOA model. This model emphasizes the integration of assessment with instruction, learner participation in evaluating performance, and the systematic use of feedback to promote forward-looking learning. Guided by this model, each 60-minute session was organized into three interrelated phases: (a) task engagement, (b) self- and peer assessment, and (c) feedback and reflection.

In the task engagement phase (20 minutes), learners completed a timed reading passage and a 10-item comprehension quiz. To scaffold metacognitive awareness, passages were annotated with margin prompts (e.g., "*What strategy helped you answer this question?*"), and key terms were bolded to draw attention to textual cues. In the assessment phase (20 minutes), students first self-rated their quiz answers using a three-point rubric (Correct = 2, Not Sure = 1, Incorrect = 0). Importantly, they were required to justify uncertain answers in their notebooks, thereby making their strategy use visible (e.g., "*I used skimming to find the main idea, but I'm not sure about the inference question*"). Next, students exchanged quizzes anonymously and applied the same rubric to their peers' responses. This activity enabled learners to negotiate criteria, compare judgments, and develop greater rating consistency. In the feedback and reflection phase (20 minutes), the teacher mediated group discussions to resolve discrepancies between self, peer, and teacher ratings. Here, feedback was not limited to correctness but extended to strategies, helping learners recalibrate their approaches for subsequent tasks.



The control group also worked with the same 18 passages and 10-item quizzes but did not receive LOA-based activities. Their sessions consisted of timed reading and comprehension practice followed by teacher-provided corrections, without opportunities for structured self- or peer assessment or collaborative reflection.

### 3.5. Data Collection Procedure

First, the data necessary for conducting this research were gathered over the summer of 2023. Prior to the intervention, all participants completed a pretest battery to establish baseline proficiency and homogeneity. This included the OPT to confirm intermediate-level proficiency, and initial self- and peer rating accuracy tasks. Pretests were administered over one session.

The experimental group participated in a 10-week intervention explicitly designed around Jones and Saville's (2016) systemic model of LOA. Their framework emphasizes that assessment should (a) function as a learning task, (b) actively involve learners in assessment processes, and (c) generate feedback that drives future learning. Each 60-minute session followed a structured cycle aligned with these principles.

Phase 1: Assessment as a learning task (15 minutes): At the start of each session, learners were presented with a 350-450-word academic passage, carefully selected for intermediate readability (Flesch-Kincaid 8.15-9.36). Passages were drawn from topics in science, humanities, technology, and social studies. Learners then completed a 5-item multiple-choice comprehension quiz under timed conditions. To support metacognitive engagement, passages included bolded key terms and margin annotations prompting reflection (e.g., "*Which clue in the passage helped you infer the answer?*"). This ensured that the quiz was not a summative test but a task embedded with opportunities for learning.

Phase 2: Self-assessment and reflection (20 minutes): After completing the quiz, learners applied the three-choice rubric to rate their own responses. Crucially, they were required to justify each rating in their notebooks, writing short comments about their reasoning (e.g., "*I scanned for dates to answer Q2*" or "*I wasn't sure about Q4 because I couldn't infer the author's attitude*"). This step encouraged strategy articulation, making tacit processes visible. The teacher circulated during this phase, clarifying rubric application, encouraging more precise justifications, and reinforcing effective strategies (e.g., "*You used context clues here effectively, how could you apply that to Q5?*"). This individualized interaction anchored the assessment within active learning.

Phase 3: Peer assessment and collaborative calibration (15 minutes): Next, quizzes were redistributed anonymously for peer assessment. Learners applied the same rubric to classmates' answers, comparing their reasoning with another learner's performance. When discrepancies arose between peer and self-ratings, learners discussed them in pairs or groups before comparing their judgments with teacher benchmarks. The teacher facilitated whole-class calibration, displaying anonymized examples (e.g., *"One peer marked this as correct, but the supporting detail was missing. Let's check paragraph 3 together"*). This step operationalized Jones and Saville's emphasis on learner involvement by making students co-constructors of assessment criteria rather than passive recipients.

Phase 4: Teacher-mediated feedback for future learning (10 minutes): Finally, the class engaged in feedback discussions guided by the teacher. Rather than providing right/wrong answers, the teacher modeled how to connect strategies to outcomes (e.g., *"Notice how summarizing each paragraph could have helped answer Q5 more efficiently"*). Feedback was delivered as *feedforward*—always linking the assessment task to future learning opportunities. Learners reflected on which strategies had worked, which needed adjustment, and how they would apply these insights in subsequent tasks.

Across all sessions, the teacher adopted a facilitative rather than directive role, scaffolding learners' reasoning, drawing attention to assessment criteria, and promoting metacognitive awareness. In contrast, the control group worked with the same passages and quizzes but received traditional teacher-led instruction. The teacher explained strategies explicitly (e.g., skimming, scanning), reviewed correct answers, and provided summative scores without engaging learners in self- or peer rating. Interaction was largely unidirectional, with fewer opportunities for reflection or collaborative dialogue. The whole procedure is illustrated in Table 1. It includes the principles of Jones and Saville's (2016) LOA classroom model and their connection with what we did in the experimental group.

The control group completed the same reading passages and quizzes over the 10-week period but followed a more traditional approach. The teacher explicitly taught reading strategies through explanation and modeling, reviewed answers aloud, and provided summative feedback (e.g., *"You scored 6/10"*). Learners practiced strategies individually and copied model responses but were not involved in self- or peer assessment activities, nor did they receive formative feedback beyond corrections.

**Table 1.** *Alignment of Classroom Procedure with Jones & Saville's (2016) LOA Model*

Jones & Saville's LOA Principle	Implementation in This Study
Assessment as a learning task	Learners completed a 350-450 word passage with a 5-item comprehension quiz designed not just for scoring, but with margin prompts and bolded keywords to stimulate strategy use and reflection.
Active learner involvement in assessment	After quizzes, learners conducted self-ratings using a 3-point rubric and wrote justifications for each response, making their thinking explicit.
Peer involvement and collaborative calibration	Quizzes were anonymously redistributed for peer-rating. Learners discussed discrepancies in small groups, followed by teacher-led whole-class calibration sessions.
Teacher role as a facilitator and guide	The teacher circulated during self/peer rating to scaffold rubric application and reinforce strategy use, adopting a facilitative rather than directive stance.
Feedback and feedforward for future learning	Whole-class discussions emphasized strategy-based feedback (e.g., summarizing, skimming, context clues). The teacher modeled how to transfer insights from current tasks to future reading challenges.
Integration of assessment and instruction	Each session followed a cycle of task → self-assessment → peer-assessment → feedback, embedding assessment within the instructional process rather than separating the two.

Posttest data collection occurred one week after the intervention. Both groups repeated the self- and peer-rating accuracy tasks, with  $\Delta$  scores calculated as in the pretests. For the experimental group, semi-structured interviews (10–15 minutes each with 10 participants) were also conducted to capture learner perceptions of LOA.

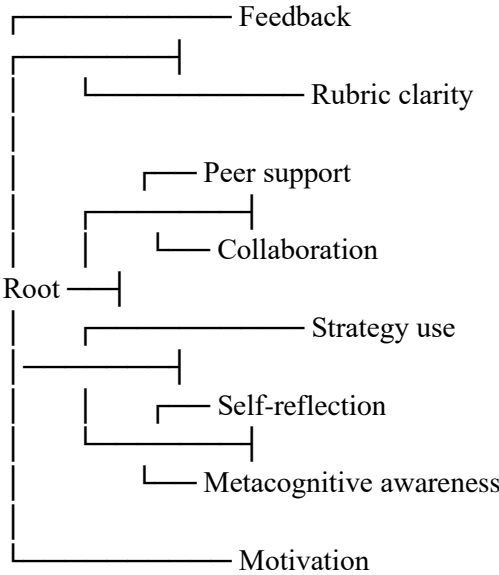
### 3.6. Data Analysis

The data analysis began with preliminary checks to ensure the assumptions of parametric tests were met. Normality was assessed using the Shapiro-Wilk test ( $p > .05$ ), and homogeneity of variance was verified via Levene's test ( $p > .05$ ). Descriptive statistics, including means and standard deviations, were calculated for all variables at pre- and posttest stages for both the experimental and control groups. The quantitative analysis employed inferential statistics to address the first four research questions. Self- and peer rating accuracy (RQ1 and RQ2) were evaluated using  $\Delta$  scores (absolute differences between learner and teacher ratings), analyzed via paired and independent t-tests to determine within- and between-group changes. All quantitative analyses, performed in SPSS

v.27, included effect size calculations (Cohen's *d*) to measure the practical significance.

The qualitative strand addressed RQ3 through thematic analysis of semi-structured interviews with 10 participants from the experimental group. Interviews were transcribed and imported into NVivo 12 for coding. Two independent coders, who had a PhD in TEFL and were experienced in coding, conducted the analysis. The intercoder agreement was calculated using Cohen's kappa ( $\kappa = .89$ ), indicating high reliability. Coding proceeded inductively in three stages: open coding identified initial units of meaning (e.g., "rubric clarity," "peer feedback value"); axial coding grouped these codes into broader categories (e.g., "metacognitive awareness," "collaborative learning"); and selective coding distilled them into overarching themes (e.g., "LOA as a catalyst for self-regulated learning"). To strengthen the dependability of findings, NVivo's cluster analysis function was employed. This tool grouped related codes based on word similarity and co-occurrence, producing a dendrogram that visually represents the relationships between themes (see Figure 2).

**Figure 2.** NVivo Cluster Analysis of Interview Data



The cluster analysis suggested that learners tended to associate feedback with rubric clarity, reinforcing the idea that transparent assessment tools functioned as a form of feedback. Similarly, peer support

clustered with collaboration, highlighting the social dimension of LOA where learners co-constructed understanding. A third cluster grouped strategy use, self-reflection, and metacognitive awareness, indicating that learners perceived reading strategies not as isolated techniques but as part of a reflective cycle of monitoring and evaluating their own learning. Finally, motivation emerged as a distinct yet connected node, suggesting that students viewed motivational gains as an outcome of the reflective and collaborative practices fostered by LOA.

## 4. Results

### 4.1. Results of Quantitative Phase

The analysis began with preliminary assumption checks: normality was confirmed via the Shapiro-Wilk test ( $p > .05$ ). Descriptive statistics (means  $M$  and standard deviations  $SD$ , minimums, and maximums) for self- and peer rating accuracy  $\Delta$  scores at pretest and posttest stages are presented in Table 2.

**Table 2.** Descriptive Statistics

Variable	Group	Time	M	SD	Min	Max
Self-Rating Accuracy ( $\Delta$ )	Experimental	Pre	8.25	3.20	2	15
Self-Rating Accuracy ( $\Delta$ )	Experimental	Post	3.58	1.89	0	8
Self-Rating Accuracy ( $\Delta$ )	Control	Pre	8.10	3.15	1	14
Self-Rating Accuracy ( $\Delta$ )	Control	Post	7.97	3.10	1	13
Peer-Rating Accuracy ( $\Delta$ )	Experimental	Pre	10.52	4.32	3	20
Peer-Rating Accuracy ( $\Delta$ )	Experimental	Post	5.14	2.05	1	10
Peer-Rating Accuracy ( $\Delta$ )	Control	Pre	10.47	4.28	2	19
Peer-Rating Accuracy ( $\Delta$ )	Control	Post	9.30	4.24	2	18
Self-Rating Accuracy ( $\Delta$ )	Experimental	Pre	8.25	3.20	2	15
Self-Rating Accuracy ( $\Delta$ )	Experimental	Post	3.58	1.89	0	8
Self-Rating Accuracy ( $\Delta$ )	Control	Pre	8.10	3.15	1	14
Self-Rating Accuracy ( $\Delta$ )	Control	Post	7.97	3.10	1	13
Peer-Rating Accuracy ( $\Delta$ )	Experimental	Pre	10.52	4.32	3	20
Peer-Rating Accuracy ( $\Delta$ )	Experimental	Post	5.14	2.05	1	10
Peer-Rating Accuracy ( $\Delta$ )	Control	Pre	10.47	4.28	2	19
Peer-Rating Accuracy ( $\Delta$ )	Control	Post	9.30	4.24	2	18

As Table 2 depicts, for self-rating accuracy ( $\Delta$ ), the experimental group showed a marked reduction in discrepancies ( $M = 3.58$ ,  $SD = 1.89$ ) post-intervention, whereas the control group's  $\Delta$  scores remained high ( $M = 7.97$ ,  $SD = 3.10$ ). Peer-rating accuracy showed this trend (Experimental:  $M = 5.14$ ,  $SD = 2.05$ ; Control:  $M = 9.30$ ,  $SD = 4.24$ ). To answer the first research question, paired samples t-tests were done on the self-rating discrepancies ( $\Delta$  scores) between the pretest and posttest (see Table 3).

**Table 3.** Paired Samples *t*-Tests for Self-rating  $\Delta$  Scores

Group	Pretest M	Posttest M	t	df	p	Cohen's d
Experimental	8.25	3.58	-7.95	27	<.001	1.71
Control	8.10	7.97	-0.21	28	.838	0.04

Table 3 depicts that the experimental group showed a statistically significant reduction in self-rating discrepancies ( $\Delta$  scores) from pretest ( $M = 8.25$ ,  $SD = 3.20$ ) to posttest ( $M = 3.58$ ,  $SD = 1.89$ ),  $t(27) = -7.95$ ,  $p < .001$ , with a large effect size ( $d = 1.71$ ). The control group exhibited no significant change ( $t(28) = -0.21$ ,  $p = .838$ ,  $d = 0.04$ ). Table 4 illustrates the results of the independent samples *t*-tests on self-rating  $\Delta$  scores between the control and experimental groups.

**Table 4.** Independent Samples *t*-Tests for Self-rating  $\Delta$  Scores

Time	Experimental M	Control M	t	df	p	Cohen's d
Pretest	8.25	8.10	0.19	55	.853	0.04
Posttest	3.58	7.97	-5.62	55	<.001	1.54

As Table 4 shows, pretest scores confirmed baseline equivalence between groups ( $t(55) = 0.19$ ,  $p = .853$ ,  $d = 0.04$ ). Posttest comparisons revealed significantly greater self-rating accuracy in the experimental group ( $M = 3.58$ ) compared to the control group ( $M = 7.97$ ),  $t(55) = -5.62$ ,  $p < .001$ ,  $d = 1.54$ . To deal with the second research question, paired samples *t*-tests were done on the peer rating discrepancies ( $\Delta$  scores) between the pretest and posttest (see Table 5).

**Table 5.** Paired Samples *t*-Tests for Peer Rating  $\Delta$  Scores

Group	Pretest M	Posttest M	t	df	p	Cohen's d
Experimental	10.52	5.14	-6.30	27	<.001	1.36
Control	10.47	9.30	-1.45	28	.162	0.30

As Table 5 displays, the experimental group demonstrated a significant improvement in peer-rating accuracy ( $\Delta$  scores) from pretest ( $M = 10.52$ ,  $SD = 4.32$ ) to posttest ( $M = 5.14$ ,  $SD = 2.05$ ),  $t(27) = -6.30$ ,  $p < .001$ ,  $d = 1.36$ . The control group, however, showed no significant change ( $t(28) = -1.45$ ,  $p = .162$ ,  $d = 0.30$ ). Table 6 demonstrates the results of the independent samples *t*-tests on peer rating  $\Delta$  scores.

**Table 6.** *Independent Samples t-Tests for Peer Rating  $\Delta$  Scores*

Time	Experimental M	Control M	t	df	p	Cohen's d
Pretest	10.52	10.47	0.04	55	.970	0.01
Posttest	5.14	9.30	-4.25	55	<.001	1.17

As illustrated in Table 6, pretest peer rating accuracy  $\Delta$  scores were equivalent between groups ( $t(55) = 0.04$ ,  $p = .970$ ,  $d = 0.01$ ). Posttest comparisons indicated significantly higher accuracy in the experimental group ( $M = 5.14$ ) compared to the control group ( $M = 9.30$ ),  $t(55) = -4.25$ ,  $p < .001$ ,  $d = 1.17$ .

#### 4.2. Results of Qualitative Phase

The semi-structured interviews with 10 experimental group participants revealed nuanced perceptions of LOA practices, particularly self- and peer rating tasks. Participants generally expressed appreciation for the reflective nature of the activities but identified challenges in consistency and clarity. Table 7 summarizes participants' responses to each interview question with response frequencies.

**Table 7.** *Summary of Participants' Responses to Interview Questions (n = 10)*

Question	Key Responses	Frequency
1. Feelings about self-rating	"Scary at first"	3/10 (30%)
	"Helped me notice my gaps"	7/10 (70%)
2. Helpfulness of peer-rating	"Learned from others' mistakes"	7/10 (70%)
	"Sometimes confusing"	3/10 (30%)
3. Difficulties in self/peer-rating	"Hard to be honest with myself"	6/10 (60%)
	"Disagreed with classmates"	4/10 (40%)
4. Usefulness of rating guides (0-2)	"Clear rules reduced stress"	7/10 (70%)
	"Wanted more examples"	3/10 (30%)
5. Suggestions for improvement	"More practice with the rubric"	6/10 (60%)
	"Group discussions first"	4/10 (40%)

Note. Participants often provided multiple responses; frequencies reflect the number of participants endorsing each theme.

As Table 7 displays, participants initially described self-rating as intimidating, with three (30%) noting it felt "*scary at first*" (P3). However, seven (70%) reported growing confidence over time, linking self-rating to improved awareness of reading strategies. For instance, P6 reflected, "*After a few weeks, I could guess which answers I got wrong before checking.*" This shift suggests that repeated engagement with LOA tasks cultivated metacognitive skills, though three struggled to reconcile self-ratings with teacher scores, leading to frustration.

Peer-rating elicited approval from seven participants (70%), who valued exposure to diverse answers. P2 explained, "*When I saw a*



*classmate's wrong answer, I thought, Oh, I almost did that!"* However, three (30%) found discrepancies between peer and teacher ratings confusing. P4 recounted, *"I marked a peer's answer as correct, but the teacher said it was wrong. I didn't understand why."* Such instances underscored the need for clearer rubrics.

The three-choice rubric (0-2 scale) was praised by seven participants (70%) for its simplicity but criticized by three (30%) for vagueness. Participants appreciated its role in *"making grading less scary"* (P7) but found the *"Not Sure"* category ambiguous. P9 remarked, *"I used 'Not Sure' when I was between correct and wrong,"* while P10 interpreted it as *"a way to ask the teacher for help."* This inconsistency highlighted the need for explicit training in rubric use.

Emotional challenges emerged as a recurring subtheme. Six participants (60%) admitted to hesitating when penalizing themselves, with P1 confessing, *"I gave myself 1 point instead of 0 because I felt bad."* Others overestimated their performance, with P7 noting, *"I thought I did better than I actually did."* These emotional barriers underscored the importance of fostering a non-judgmental environment for self-assessment.

Suggestions for improvement centered on scaffolding and collaboration. Six participants (60%) advocated for *"practice sessions with sample answers"* (P8), and four (40%) requested *"group discussions before rating"* (P5) to align rubric interpretations. P8 emphasized, *"If we talked about what good answers look like, peer-rating would be fairer."* These recommendations align with the theme of structured guidance, emphasizing preparatory activities to reduce ambiguity. Table 8 outlines the emergent themes with illustrative quotes.

As presented in Table 8, participants recognized that self-assessment encouraged reflection on their thought processes, helping them identify gaps in understanding and track progress over time. For example, P6 reflected, *"Rating myself made me think: Why did I choose this answer?"* This metacognitive awareness was a recurring benefit, as learners began to question their reasoning during self-rating, fostering critical thinking and strategy awareness. Over time, participants linked self-rating to improved self-regulation, such as P6's ability to predict errors before teacher feedback. However, challenges arose when discrepancies between self- and teacher ratings caused confusion, prompting deeper reflection. These findings suggest that self-rating activities can cultivate metacognitive skills but require scaffolding, such as post-rating discussions, to resolve uncertainties and reinforce learning.

**Table 8. Emergent Themes and Descriptions**

Theme	Description	Example Quote
Self-Rating as a Tool for Metacognitive Growth	Participants recognized self-assessment as a means to identify weaknesses and track progress.	“Rating myself made me think: Why did I choose this answer?” (P6).
Peer Rating as a Collaborative Learning Opportunity	Peer evaluation fostered collective understanding but caused tension when ratings diverged.	“Seeing others’ mistakes helped me avoid them” (P2).
Ambiguity in Rubric Application	While the 0-2 scale was generally clear, some struggled with interpreting “Not Sure” (1 point).	“Is ‘Not Sure’ for me or the teacher?” (P7).
Emotional Barriers to Honest Self-Assessment	Fear of harsh self-judgment or overconfidence affected accuracy.	“I didn’t want to admit I was wrong” (P1).
Desire for Structured Guidance	Participants requested more examples and preparatory activities.	“Show us how to use the rubric before quizzes” (P8).

Peer evaluation fostered collective learning, as participants analyzed classmates’ answers to refine their own strategies. P2 explained, “*Seeing others’ mistakes helped me avoid them,*” highlighting how peer-rating exposed learners to diverse problem-solving approaches and enhanced critical analysis. However, tensions emerged when peer ratings conflicted with teacher evaluations, underscoring the need for rubric clarity. Some participants also reported increased empathy during peer-rating, as the activity normalized mistakes and reduced fear of judgment. For instance, P2 shared, “*I started being kinder because I knew how it felt to be graded.*” These insights suggest that peer-rating promotes collaborative learning but requires structured mediation, such as teacher-led calibration, to align interpretations and ensure fairness.

While the 0-2 scale simplified scoring, participants struggled with inconsistent interpretations of the “Not Sure” (1-point) category. P7 asked, “*Is ‘Not Sure’ for me or the teacher?*” reflecting widespread confusion. Some learners used “Not Sure” to flag personal uncertainty (e.g., P5), while others saw it as a signal for teacher help (e.g., P4). This variability led to mismatches between learner and teacher ratings, undermining rubric reliability. Participants also desired concrete examples, such as annotated responses, to standardize rubric use. These findings highlight the importance of balancing simplicity with explicit criteria in rubric design, particularly for intermediate learners who may

lack the linguistic or cognitive maturity to interpret nuanced scoring systems.

Fear of self-criticism or overconfidence hindered accurate self-rating. P1 admitted, "*I didn't want to admit I was wrong*," illustrating how emotional resistance affected self-assessment. Participants often avoided harsh self-penalties, such as awarding 1 point instead of 0 for incorrect answers. Others overestimated their performance, attributing errors to "*bad luck*" rather than knowledge gaps. These emotional barriers underscore the need to normalize mistakes and create a non-judgmental environment for self-assessment. Teacher modeling, such as sharing anonymized errors, could help reduce resistance and encourage honest reflection. Participants emphasized the need for preparatory training and collaborative calibration to improve LOA activities. P8 suggested, "*Show us how to use the rubric before quizzes*," reflecting a common request for rubric familiarization. Additionally, participants advocated for post-rating debriefs to address lingering doubts and reinforce learning. These suggestions highlight the importance of scaffolding LOA activities with pre- and post-rating support mechanisms.

## 5. Discussion

Regarding the first research question, the findings provided compelling evidence for the positive effect of the LOA intervention on EFL learners' self-rating accuracy. The statistically significant reduction in self-rating discrepancies observed in the experimental group, verified by a large effect size (Cohen's  $d = 1.71$ ), indicates a significant enhancement in participants' metacognitive adjustment. This finding validates the core LOA principle that involving students in the assessment process serves as a powerful factor for the development of self-regulatory capabilities (Carless, 2007, 2015; Yan & Carless, 2022). The participants' transition from passive recipients of assessment to active agents of their own learning represents the concept of assessment as a mediational tool for cognitive development, a central principle of sociocultural theory (Vygotsky, 1978; Tsagari & Banerjee, 2016). This progress aligns with theoretical propositions that LOA fosters greater learner autonomy and ownership over the learning process (Al-Abri et al., 2025). However, the qualitative data revealed significant affective barriers, such as the reluctance to engage in honest self-penalization, which potentially reduces rating accuracy. This observation underscores the assertion by Beikmohammadi et al. (2020) that the efficacy of self-assessment is dependent not only on the provision of criteria but also on the creation of a supportive pedagogical environment that alleviates assessment anxiety.

Consequently, these findings suggest that the development of metacognitive accuracy is associated with the management of affective factors.

Turning to the second research question, this study also confirmed that LOA enhanced peer rating accuracy. The statistically significant improvement in the experimental group, with a large effect size (Cohen's  $d = 1.36$ ), underscores that peer assessment, when embedded in a guided LOA cycle, can foster collaborative construction of knowledge. Peer rating requires learners to apply assessment criteria to external performances, a process that, as Topping (2010) argued, stimulates evaluative judgment and deepens understanding of quality standards. The findings are also in line with Liu and Carless (2006), who emphasized the value of peer assessment in promoting dialogic engagement and collective responsibility for learning. However, the data suggested that peer rating posed more difficulties than self-rating, as reflected in consistently higher mean discrepancy scores. This finding is not surprising: evaluating the work of another learner involves additional interpretive and social challenges, such as deciphering unfamiliar reasoning and negotiating interpersonal dynamics (Al-Rashidi et al. (2022). The qualitative findings confirmed these complexities, with participants reporting frequent disagreements about rubric interpretation, particularly concerning the "Not Sure" category. Such interpretive variation reflects what Sadler (2010) calls "the problem of tacit criteria," whereby learners may lack a shared conception of quality despite having access to explicit descriptors.

The third research question probed learner perceptions of LOA, which were largely positive despite acknowledged challenges. Learners valued self-rating as a driver of metacognitive growth and peer assessment as a collaborative learning opportunity, reflecting Carless's (2015) observation that LOA fosters deeper engagement when learners understand its rationale. Al-Rashidi et al. (2022) similarly reported that integrating self-, peer-, and teacher assessments enhanced learners' performance and attitudes toward assessment tasks. Conversely, the perceptual data also served to identify salient obstacles to optimal implementation. Challenges such as Ambiguity in Rubric Application and Emotional Barriers provide empirical corroboration for the practical difficulties documented in previous research, particularly in contexts characterized by a strong summative assessment culture (Khalili et al., 2024). The observed ambiguity surrounding the Not Sure category in the scoring rubric illustrates a fundamental principle of assessment design: that criteria must not only be explicit but also be subject to communal negotiation and exemplification to ensure consistent interpretation.

Furthermore, the participants' constructive suggestions for more practice with the rubric and group discussions first themselves constitute evidence of developing assessment literacy. These recommendations provide a clear directive for educators to prioritize preparatory calibration activities and to view the LOA framework as a dynamic system requiring repeated refinement based on learner feedback, thereby fully embodying the feed-forward principle essential to this approach (Carless, 2007; Dehghanpour et al., 2022; Salajegheh et al., 2022).

## 6. Conclusion

This study aimed to explore the implementation of LOA, focusing on the role of self- and peer assessment in enhancing Iranian EFL learners' rating accuracy and perceptions of assessment. The findings indicated that LOA, when systematically embedded into classroom practice, can significantly improve learners' assessment judgment. Specifically, the use of simple, structured rubrics and teacher-mediated discussions enabled learners to make more accurate self-ratings, thereby confirming that self-assessment is not inherently unreliable but can be a powerful tool for metacognitive development when scaffolded appropriately (Carless, 2007, 2015; Yan & Carless, 2022). Similarly, peer assessment yielded significant benefits, as learners who engaged in guided peer-rating activities demonstrated improved accuracy over time. These results align with the principles of LOA, which emphasize the integration of assessment, teaching, and learning in a coherent cycle (Jones & Saville, 2016).

This study contributes to the growing body of evidence that LOA provides learners with more than opportunities to demonstrate knowledge: it actively fosters reflection, collaboration, and assessment literacy. By shifting learners from passive recipients of scores to active participants in the evaluative process, LOA creates conditions for the development of self-regulated learning (Tsagari & Banerjee, 2016; Al-Rashidi et al., 2022). Importantly, the findings underscore that both cognitive and affective factors shape the success of LOA interventions. While rubrics and teacher guidance improved rating accuracy, learners still reported hesitation in penalizing themselves and occasional ambiguity in applying peer-assessment criteria. These insights suggest that the effectiveness of LOA is contingent upon continuous scaffolding and a supportive classroom climate.

The present study is not without limitations. The sample was restricted to intermediate-level Persian-speaking EFL learners at a single private institute in Tehran, which limits the generalizability of the findings to

other proficiency levels, institutional settings, and cultural contexts. Moreover, the intervention was relatively short in duration, raising questions about the sustainability of the observed gains. Future research should, therefore, replicate this study in varied contexts and extend the intervention period to examine long-term effects. In addition, further work could investigate how LOA might operate in digital or mobile-assisted learning environments (Al-Abri et al., 2025) and whether extended calibration activities can further reduce interpretive discrepancies in peer assessment. Comparative studies across proficiency levels and age groups would also enrich our understanding of the adaptability of LOA practices.

**Funding:** This research received no external funding from any agency.

**Conflicts of Interest:** The authors declare no conflict of interest.

## 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.
- Al-Rashidi, A. H., Asif, M., Vanani, M. G., & Aberash, A. (2022). Learner-oriented assessment (LOA) practice: the comparative study of self-assessment, peer-assessment, and teacher assessment on EFL learners' writing complicity, accuracy, and fluency (CAF), speaking CAF, and attitude. *Language Testing in Asia*, 12(1), 1–22. <https://doi.org/10.1186/s40468-022-00209-x>
- 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>
- 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. (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>



- Dehghanpour, B., Beheshti, Z., & Dehestani, M. (2022). The role of corrective feedback via offline and online applications on Iranian EFL learners' grammatical accuracy. *Journal of Language, Culture, and Translation*, 4(2), 169–183. <https://doi.org/10.30495/lct.2022.1972989.1074>
- Educational Testing Service. (2012). *The official guide to the TOEFL iBT test* (6th ed.). McGraw-Hill Education.
- 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>
- Hume, A., & Coll, R. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education*, 16(3), 269–290. <https://doi.org/10.1080/09695940903319661>
- 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>
- Khodami, M. (2023). Association between emotional intelligence and meta-cognitive reading strategies among Iranian intermediate EFL learners. *Journal of Language, Culture, and Translation*, 5(2), 250–266. <https://doi.org/10.71864/lct-2024-1130184>
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer-assessment. *Teaching in Higher Education*, 11(3), 279–290. <https://doi.org/10.1080/13562510600680582>
- Nazari, A. (2023). Impact of using the STAD model of cooperative learning on autonomy and attitude of Iranian secondary school EFL learners. *Journal of Language, Culture, and Translation*, 5(2), 48–76. <https://doi.org/10.30495/lct.2023.1982919.1085>
- Panadero, E., Brown, G., & Strijbos, J.W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803–830. <https://doi.org/10.1007/s10648-015-9350-2>
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge University Press.
- Salajegheh, S., Khomeijani Farahani, A., & Shahabi. H. (2022). The role of explicit corrective feedback timing in second language structure accuracy. *Journal of Language, Culture, and Translation*, 4(2), 1–21. <https://doi.org/10.30495/lct.2022.1947452.1050>



- Topping, K. (2017). Peer-assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, 1(1), 1–17. <https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007>
- Tsagari, D., & Banerjee, J. (Eds.). (2016). *Handbook of second language assessment* (Vol. 12). Walter de Gruyter GmbH & Co KG.
- 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>
- Zeinali, E, Afraz, S., & Samimi, F. (2025). Construction and development of a questionnaire of grammar learning strategies with a focus on the most and least utilized strategies: Insights from Iranian EFL learners. *Journal of Language, Culture, and Translation*, 8(1), 1–26. <https://doi.org/10.71864/lct-2025-1208144>

## Appendices

### Appendix 1. Samples of Self-Rating and Peer Rating Tasks

**Why Coral Reefs Are Important**

Coral reefs are colorful underwater structures found in warm, shallow ocean waters. They are made by tiny animals called coral polyps. These polyps build hard shells around themselves. When many of them live together, they form large reefs. Over many years, these reefs become big homes for fish, sea turtles, and other sea creatures.

Coral reefs are sometimes called the "rainforests of the sea" because they have so many different animals living in them. Although they cover only a small part of the ocean floor, about 25% of ocean animals depend on them for food or shelter.

Coral reefs are also helpful to people. They protect coastlines from big waves and storms by slowing down the water before it reaches the land. They are important for local businesses too. Many people earn money from fishing and tourism near coral reefs. Tourists come to swim, dive, and look at the beautiful sea life.

Coral reefs also have a special connection with tiny plants called algae. These algae live inside the coral and give the coral food through a process called photosynthesis, which uses sunlight. In return, the coral gives the algae a safe place to live. This helps both the coral and the algae grow strong and healthy.

Sadly, coral reefs are in danger. When ocean water becomes too warm, corals get stressed and push out the algae living inside them. This is called coral bleaching. Without the algae, the coral loses its color and its main food source. If the water stays warm too long, the coral can die.

Other problems include pollution, overfishing, and people damaging reefs by walking or boating on them. Scientists and governments are working to protect coral reefs. They are creating special areas where fishing is not allowed and teaching people how to care for the ocean. Some scientists are even growing new coral in labs to help fix damaged reefs.

1. What creates the hard structure of coral reefs?

A. Fish bones  
B. Rocks and sand  
✓ C. Coral polyps  
D. Sea plants

2. Why are coral reefs called the "rainforests of the sea"?

A. They are hot and wet  
B. They grow fast  
✓ C. They have many animals living in them  
D. They are hard to find

3. How do coral reefs help people?

A. They clean drinking water  
✓ B. They protect land from storms  
C. They grow seaweed for food  
D. They create fresh water

4. What happens during coral bleaching?

✓ A. Corals lose their algae and color  
B. Corals grow faster  
C. Corals turn green from algae  
D. Fish leave the coral reef

5. What is one way people are helping coral reefs?

A. Building houses near them  
✓ B. Making special protected areas  
C. Throwing food into the ocean  
D. Catching more fish

Q	Student's Score (A)	Student's Score (B)	Teacher's Score
1	2	2	2
2	1	1	2
3	1	2	2
4	0	1	2
5	1	0	2
Total	5	6	10

Self-rating average:  
5 = 10 ÷ 5 = 2

Peer-rating average:  
6 = 10 ÷ 4 = 2.5

## The Effect of Learning-Oriented Assessment on EFL Learners' Peer and Self-Rating Accuracy ...

### The Printing Press and How It Changed the World

A long time ago, before machines were invented, books had to be written by hand. This took a lot of time, sometimes many months for just one book. The people who copied books were often monks in churches. Because it was such slow work, books were very expensive. Only rich people, churches, or kings could own them. Most people never saw a book in their lives, and many could not read or write.

This changed in the 1400s when a man named Johannes Gutenberg invented a new way to make books. He lived in Germany and created a machine called the printing press. It used small metal letters that could be moved and reused. These letters were put into a frame to form words and pages. Then, ink was added and paper was pressed on top to print the page.

Gutenberg's press made it possible to print many books quickly and cheaply. One of the first big books he printed was the Bible. After that, printing spread to other countries in Europe. More books were made, and more people began to learn to read.

The printing press changed the world in many ways. Before, only a few people had knowledge. Now, books and ideas could travel far and fast. People could learn about science, art, religion, and other cultures. New thinking helped start big changes in history, like the Renaissance and the Reformation.

Not everyone was happy about this. Some kings and religious leaders were afraid people would read new ideas and question old beliefs. They tried to stop certain books, but they could not stop the change.

Today, we use modern tools like the internet, but it all started with Gutenberg's invention. The printing press helped more people read, think, and share ideas. It is one of the most important inventions in human history.

- Why were books very expensive before the printing press was invented?
  - They were made of gold
  - They had many pictures
  - Only kings were allowed to own them
  - ☒ They were written by hand and took a long time to copy
- What was special about Gutenberg's printing press?
  - It used paper made from trees
  - ☒ It could write books by itself
  - It used small metal letters that could be reused
  - It made handwritten copies faster
- What was one of the first important books printed by Gutenberg?
  - The dictionary
  - ☒ The Gutenberg Bible
  - A science textbook
  - A storybook for children
- How did the printing press help more people learn?
  - ☒ It made books cheaper and easier to find
  - It taught people how to speak Latin
  - It helped people find jobs
  - It made people travel more
- Why did some leaders not like the printing press?
  - It used too much ink
  - ☒ It spread new ideas that could challenge their power
  - It made people stop going to church
  - ☒ It caused people to forget old languages

Q	Student's Score (A)	Student's Score (B)	Teacher's Score
1	2	2	2
2	1	1	0
3	0	0	2
4	1	2	0
5	2	2	0
Total	6	7	4

Self-rating accuracy:

$$6 - 4 = +2$$

Peer-rating accuracy:

$$7 - 4 = +3$$