Innovation in Language Testing and Teaching (ILTT)
Volume 2 Issue 1 Year 2026 Pages 95 - 118
E–ISSN : 3110-5483 | DOI: 10.61227
E-mail : iltt.pusmedia@gmail.com
https://journal.pusmedia.com/index.php/ILTT/index

Exploring the Differential Impact of Self-Assessment, Peer-Assessment, and Instructor-Assessment on the Enhancement of EFL Writing Proficiency

Fatemeh Kheybarzadeh Tizkharab [1]*

[1] Islamic Azad University of Urmia, Urmia, Iran

Correspondence e-mail * : Kheybarzadehfatemeh@gmail.com

Abstract : The present study employed a mixed-method design to investigate the effectiveness of alternative assessment in enhancing language performance. A total of 70 Iranian English as a Foreign Language (EFL) learners participated in this research. All participants were advanced students enrolled in three intact classes at an online language center located in Urmia, Iran. The primary focus of the study was to examine the impact of self-assessment and peer-assessment in comparison with instructor-assessment on the development of writing skills among Iranian EFL learners within an online learning environment. For the quantitative phase of the study, which followed a true-experimental design, the three intact classes were randomly assigned into two experimental groups (self-assessment group, N=25; peer-assessment group, N=25) and one control group (instructor-assessment group, N=20). To address the research questions of the quantitative phase, a mixed-design Analysis of Covariance (ANCOVA) was conducted. This analysis included one within-subject variable (writing performance), one between-subject variable (group type), and two covariates (scores from the Oxford Online Placement Test [OOPT] and the writing pre-test). The findings indicated that the self-assessment group achieved significantly higher post-test scores in both rating measures compared to the peer-assessment and instructor-assessment groups. Although all three groups demonstrated significant improvement from pre-test to post-test, the self-assessment group exhibited the greatest gains. The qualitative phase of the study, which was exploratory, descriptive, and interpretive in nature, revealed that the majority of students believed that engaging in self-assessment processes—when supported by instructor guidance—contributed positively to their writing development. Additionally, the results showed that students were more capable of identifying their writing difficulties and improving their performance when they received feedback from their peers.

Keywords: Assessment, Instructor-assessment, Online Assessment, Peer-assessment, Self Assessment.

Article info: Submitted : 2026-05-24 | Accepted : 2026-06-08 | Published : 2026-06-08

Copyright © 2026, Authors.

This is an open-access article under the CC BY 4.0

INTRODUCTION

Despite the growing recognition of alternative assessment in language education, empirical research on self- and peer-assessment specifically within online writing

contexts remains remarkably scarce as of 2026 (Panadero, et al, 2026). Recent scholarly reviews have consistently called for more rigorous mixed-methods investigations into how these formative practices operate in virtual environments, where reduced learner-instructor interaction and emerging technological challenges (such as AI-generated content) significantly affect feedback quality and implementation fidelity.

Assessment is a fundamental component of language instruction, playing a critical role in evaluating the effectiveness of teaching methodologies, monitoring learner progress, and making informed pedagogical decisions (Tsagari, 2017). The significance of this issue is further heightened in online learning environments, where reduced student interaction poses additional challenges (Tartavulea et al., 2020). In response, language teachers employ a variety of assessment methods, and researchers have proposed numerous approaches to evaluate student progress (Vogt & Tsagari, 2014). These methods span a continuum from summative approaches (primarily instructor-assessment) to formative strategies (including self- and peer-assessment). All such methods are utilized to assess language skills, among which writing assessment holds particular importance, as writing proficiency is widely regarded as a hallmark of educated individuals (Deane, 2011).

Furthermore, the assessment of writing skills has become increasingly crucial in contemporary contexts, given that much of today's significant communication and correspondence occurs through emails and formal letters (Tsagari, 2017). Consequently, educators and assessors encounter increasingly complex texts (Tsagari, 2017) that demand efficient and reliable assessment methods. Therefore, existing assessment approaches must be critically examined and compared to identify the most suitable method for specific tasks. Assessing writing in online learning environments presents a particularly important challenge, as the transition from face-to-face to online assessment remains a novel endeavor for many English instructors (Anasse & Rhandy, 2021).

Generally speaking, limited attention has been devoted to online assessment practices and the influencing factors within this domain (Zou, Kong & Lee, 2021). Accordingly, further research is required in this area to address the needs of both teachers and learners, particularly given the numerous challenges that have emerged in this field. More importantly, instructors face specific obstacles related to ensuring the originality of students' writing (i.e., plagiarism detection) and managing time effectively when conducting writing assessments in online learning contexts (Dwiyanti & Suwastini, 2021).

Meanwhile, learners are less likely to respond consistently to all feedback they receive and may experience discomfort or frustration when exposed to continuous teacher feedback (Jonsson, 2013). Given the limited time typically allocated to L2 writing courses, instructors may find it difficult to provide feedback on multiple drafts of

learners' work, especially in relatively large classes. In light of these seemingly justified obstacles faced by L2 writing practitioners, many writing researchers have emphasized self-assessment and peer-assessment as suitable and engaging alternatives to traditional teacher-fronted writing assessment courses (Jensen & Fischer, 2005; Topping, 1998).

In online settings, numerous challenges related to writing assessment persist, including the need for computer integration training for writing instructors, a lack of computer-based writing instruction, and teachers' reluctance to migrate face-to-face activities to virtual spaces. Moreover, since the outbreak of the pandemic, writing assessment has increasingly demanded digital writing skills and foundational technological competencies (Anasse & Rhandy, 2021). To this end, the present study sought to explore the effectiveness of alternative assessment approaches (i.e., self- and peer-assessment) in enhancing language performance. In this regard, it focused on examining the effect of self- and peer-assessment versus instructor-assessment on improving the writing skills of Iranian EFL learners in online contexts by addressing the following research questions:

1. Does a statistically significant difference exist between the writing performance of learners assessed by their instructor and that of learners engaged in self- and peer-assessment processes?
2. Among the three assessment modalities—instructor-assessment, peer-assessment, and self-assessment—which one proves most effective in promoting writing achievement in an online EFL context?
3. Are the variations in assessment methods (self-, peer-, and instructor-assessment) systematically associated with corresponding differences in students' final examination outcomes?

Review of the Related Literature

A comprehensive meta-synthesis of EFL writing assessments from 2018 to 2024 confirms a wide variety of practices, including instructor, peer, and self-assessment, while emphasizing the necessity for more standardized and contextually valid approaches tailored to specific learning environments .According to Harlen (2005), summative assessment in language education refers to the documentation or reporting of learners' academic achievements. In the context of summative assessment, the data obtained through assessment serve as an indicator and contribute to decision-making processes; therefore, the validity and reliability of such assessments are of paramount importance (Wil Meeus, Peter Van Petegem & Nadine Engels, 2009). In essence, this form of assessment seeks to determine what the learner has acquired over a past period of instruction.

Furthermore, Mosquera, Macías, and Fernando (2015) asserted that summative assessment is an evaluative approach aimed at informing instructors about a learner's level of success based on a numerical scale. In a related study, Qu and Zhang (2013) explored the respective functions of summative and formative assessment within the college English assessment system.

On the other hand, Spiller (2012) argued that through peer-assessment, learners evaluate the quality of their peers' work or performance based on predefined criteria of excellence. In peer-assessment, students make judgments regarding the work of other students. This method is commonly employed in the evaluation of projects and practical tasks. Similarly, peer-assessment may be applied at various points throughout the learning process, providing continuous engagement for evaluators and ongoing feedback on progress to those being assessed (McDonald & Savin-Baden, 2004). In addition, this process enhances learners' metacognitive skills, and the development of metacognitive abilities subsequently facilitates self-assessment, enabling students to become more aware of what to learn, how to learn, and how to evaluate the effectiveness of their learning (Lee & Mak, 2018).

Meanwhile, during self-assessment, the learner initiates the learning process by setting goals, monitoring progress, regulating cognition, behavior, and motivation, all guided by established objectives and contextual factors (Panadero, Brown, & Strijbos, 2016). It is commonly held that in self-assessment, the learner employs metacognitive strategies to evaluate their own abilities and progress (Rivers, 2001). As noted earlier, peer-assessment is a process in which individuals within the same group evaluate one another based on specific criteria (Falchikov, 1995). The primary objective of peer-assessment is to actively engage learners in the evaluation of their own progress, thereby enabling them to enhance their learning through assessment (Orsmond, 2004).

Moreover, Birjandi and Siyyari (2011) demonstrated that while both self-assessment and peer-assessment practices contribute to improving writing performance and the accuracy of learner ratings, the peer-assessment group outperformed the self-assessment group in terms of overall writing achievement. Additionally, Ratminingsih, Marhaeni, and Vigayanti (2018) found that self-assessment is effective in enhancing learner autonomy and writing ability. Nevertheless, the findings of Ebadi and Rahimi's (2019) study revealed that learners hold positive perceptions regarding the influence of online dynamic assessment (DA) on the development of academic writing skills. In a different vein, Akimov and Malin (2020) showed that oral examinations can assist learners in overcoming some of the challenges associated with online learning.

In a similar manner, researchers and scholars in the field of writing assessment (e.g., Urbach, 2014) have proposed various models for evaluating writing tasks, noting

that methods of writing assessment vary depending on the context and the type of assessment being conducted. The effectiveness of writing tasks can be enhanced through the use of self-assessment and peer-assessment (Mowl & Pain, 1995). Finally, Fahimi, Rahimi, and Shams (2013) inv estigated the relationship between self-assessment in writing and self-regulation among Iranian EFL learners.

METHODOLOGY

Participants and Sampling

To achieve the objectives of this study, a total of 70 Iranian EFL learners were recruited as participants. These participants were advanced-level students enrolled in three intact classes at an online language center located in Urmia, Iran. Their proficiency level was confirmed by their performance on the Oxford Online Placement Test (OOPT), which classified all of them as advanced learners. The participants comprised both male and female students, ranging in age from 16 to 25 years, with a mean age of 23.5. The three intact classes were randomly assigned into two experimental groups (n = 25 and n = 25) and one control group (n = 20). The classes were randomly designated as the self-assessment, peer-assessment, and instructor-assessment groups, respectively. At the time of the study, all students were engaged in learning English writing as a compulsory component of their course curriculum.

Materials and Instruments

The following instruments were used in phase one; the quantitative (true experimental) design: Language Proficiency Test, Timed-writing Essays (pretest and posttest), and The Writing Scoring Scale. In order to homogenize the participants in terms of general English proficiency, Oxford Online Placement Test (OOPT) was administered to the students of the three groups (self-, peer- and instructor-assessment). OOPT is considered as a reliable and valid standard English proficiency test which can be administered to different number of learners with various proficiency levels (Allen, 2004). The version of OOPT that was used in this study was online and contains multiple-choice items measuring vocabulary, grammar, and reading. There was also an optional writing section.

In addition, to measure the writing performance of the participants before (i.e., as the pretest) and after the treatment (i.e., as the posttest), two 45-minute writing essays were administered to the participants of the three groups. To this end, two general topics which do not require any particular technical background knowledge was assigned for each administration. It should be noted that students took part in the test online using adobe connect, so it was possible for the instructor to test them during a special time limit.

Moreover, in order to score the participants' essays, Jacobs et al.'s (1981) writing scale was used which is mainly based on an analytical scoring procedure. The effectiveness of this writing scoring rubric has been verified by Brown and Baily (1984). According to this scale, a written text or an essay should be evaluated against a set of five criteria or subcategories such as content, organization, vocabulary, language use, and mechanics. The given scores of the two raters as well as those of the researchers were evaluated by Cohen's Kappa's interrater reliability test (Livingston, S. A., 2018).

In the phase two, the following instruments were used as the qualitative design: Students' self-assessment of Pre-test and Post-test, Students' Final self-assessment of the Whole Process, and Semi-structured interview questions. At the end of the course, students took an institutional final exam at the end of each level. Later, students filled out a final self-assessment form of their communicative competence, plans to improve, and whether or not they considered they had passed the class objectives. The form consisted of eleven parts designed by the researcher. Finally, semi-structured interview questions were used to assess students' views of the process.

Procedure

Prior to the commencement of the experiment, the Oxford Online Placement Test (OOPT) was administered to the participants of the three intact classes to ensure their homogeneity in terms of language proficiency. Subsequently, a timed writing task was given to all three groups (self-assessment, peer-assessment, and instructor-assessment) as the pre-test of the study in order to measure their writing performance before the interventions began. It should be noted that the students completed the test online using the Adobe Connect platform, which enabled the instructor to administer the test within a specified time limit.

The instruction, which consisted of an online writing course, was delivered as a single two-hour session per week over a period of 20 weeks via the Adobe Connect platform. Throughout the 20-session course, the fundamentals of paragraph writing were taught. To fulfill the objectives of the present study, identical materials and the same course book were used by the same instructor (the researcher) across all three classes. Moreover, the instructor-assessment group (control group), the self-assessment group, and the peer-assessment group (both experimental groups) received their respective assessment practices. In addition, students in the self- and peer-assessment groups received training on how to utilize a writing scoring rubric to evaluate their own written tasks and essays as well as those of their peers.

Upon completion of the online treatment sessions, another timed writing task was administered to all three groups as the post-test to measure their writing ability. As with

the pre-test, students took the post-test online using the Adobe Connect platform, allowing the instructor to conduct the test within a designated time limit.

For the second phase of the study, the researcher conducted three online conferences aimed at exploring participants' perceptions of the assessment process (self-, peer-, or instructor-assessment) during the online course, as well as their insights regarding the impact of the implementation before, during, and after the course.

At the conclusion of the online course, students took an institutional final examination at the end of the term. Subsequently, students completed a final self-assessment form addressing their communicative competence, plans for improvement, whether they believed they had achieved the course objectives, and their insights regarding the implementation of the specific type of assessment they had experienced (self-, peer-, or instructor-assessment). The interviews were conducted in Farsi, and quotations from these interviews were translated into English for the purposes of this study. Finally, the researcher conducted another online conference aimed at exploring students' perceptions of their learning process during the online course and their insights regarding the impact of implementing the three different types of assessment.

Data Analysis

1. Quantitative Data Analysis

To address the quantitative research questions, a three-step analytical procedure was followed. First, in order to verify the homogeneity of the three groups prior to the intervention, an independent-samples t-test was performed to compare the mean Oxford Online Placement Test (OOPT) scores of the self-assessment, peer-assessment, and instructor-assessment groups. Second, to investigate the effectiveness of the alternative assessment interventions, two paired-samples t-tests were conducted separately for the self-assessment and peer-assessment groups. These analyses were designed to evaluate the significance of the change in writing performance from the pre-test to the post-test within each experimental group. Third, a one-way Analysis of Covariance (ANCOVA) was conducted on the post-test writing scores, with the pre-test scores and OOPT scores serving as covariates. This analysis aimed to determine whether significant differences existed among the three groups (self-assessment, peer-assessment, and instructor-assessment) in terms of their post-treatment writing performance.

2. Qualitative Data Analysis

For the qualitative phase, the analysis of the transcribed interview data integrated both inductive and deductive coding approaches. Specifically, the presence or absence of predetermined thematic categories was examined within the participants' perceptions, while simultaneously allowing emergent themes to arise from the data.

RESULTS AND DISCUSSION

Results

The allocation of participants to each experimental condition is summarized in Table 1. According to Table 1, the instructor-assessment group contained 20 participants, whereas the self-assessment and peer-assessment groups each included 25 participants

Table 1. The Frequency Distributions of the Three Groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Instructor-Assessment Group	20	30.7	30.7	30.7
	Peer-Assessment Group	25	34.7	34.7	68.3
	Self-Assessment Group	25	36.7	34.7	100.0
	Total	70	100.0	100.0	

Table 2 provides the descriptive statistics for the three groups' performance on the Oxford Online Placement Test (OOPT), the writing pre-test, and the writing post-test. It should be noted that all pre-test and post-test writing samples were evaluated by two independent raters.

Table 2. Descriptive Statistics of the Variables by Group

Group		N	Range	Min	Max	Mean	SD	Variance
Instructor-Assessment Group	OOPT	20	18.00	80.00	98.00	87.0909	5.57980	31.134
	Pre-test Rater 1	20	16.00	60.00	76.00	67.0000	4.15188	17.238
	Pre-test Rater 2	20	15.00	61.00	76.00	67.0909	4.03448	16.277
	Post-test Rater 1	20	16.00	80.00	96.00	90.5000	3.46066	11.976
	Post-test Rater 2	20	15.00	81.00	96.00	90.5455	3.33420	11.117
	Valid N (listwise)	20						
Peer-Assessment Group	OOPT	25	16.00	80.00	96.00	89.0526	5.23316	27.386
	Pre-test Rater 1	25	19.00	60.00	79.00	72.4737	5.43004	29.485
	Pre-test Rater 2	25	19.00	60.00	79.00	72.2632	5.48576	30.094
	Post-test Rater 1	25	13.00	85.00	98.00	91.1053	3.54173	12.544

	Post-test Rater 2	25	13.00	85.00	98.00	91.1053	3.41394	11.655
	Valid N (listwise)	25						
Self-Assessment Group	OOPT	25	16.00	82.00	98.00	92.3684	4.39963	19.357
	Pre-test Rater 1	25	8.00	73.00	81.00	77.5263	2.52473	6.374
	Pre-test Rater 2	25	7.00	73.00	80.00	77.2105	2.34708	5.509
	Post-test Rater 1	25	8.00	90.00	98.00	94.7895	2.12339	4.509
	Post-test Rater 2	25	6.00	92.00	98.00	94.8421	1.83373	3.363
	Valid N (listwise)	25						

Note: OOPT: Oxford Online Placement Test

OOPT Scores

For the entire sample of 70 participants, OOPT scores ranged from a minimum of 80 to a maximum of 98, with a mean score of 89.38 (SD = 5.49), as displayed in Table 2.

Pre-test Writing Scores (by Rater 1 and Rater 2)

On the writing pre-test, Rater 1 awarded scores between 60 and 81 (M = 72.06, SD = 6.02). Rater 2, meanwhile, assigned scores ranging from 60 to 80 (M = 71.93, SD = 5.86).

Post-test Writing Scores (by Rater 1 and Rater 2)

On the writing post-test, scores given by Rater 1 ranged from 80 to 98 (M = 92.05, SD = 3.61). Similarly, Rater 2 assigned scores between 81 and 98 (M = 92.08, SD = 3.48).

1. Tests of Normality

The normality of the data distribution was examined using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The results of these normality assessments are summarized in Table 3.

Table 3. Tests of Normality

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
OOPL	.145	70	.203	.931	60	.202
Pre-test Rater 1	.137	70	.107	.932	60	.103
Pre-test Rater 2	.149	70	.102	.926	60	.101

Post-test Rater 1	.122	70	.057	.950	60	.056
Post-test Rater 2	.125	70	.060	.955	60	.066
a. Lilliefors Significance Correction						

a. Normality Assumption

The results revealed no statistically significant departures from normality in the data distribution. Consequently, the assumption of normality was met, and the application of parametric statistical procedures was deemed appropriate.

2. Inter-rater Reliability

The reliability of the scores assigned by the two raters was examined separately for the pre-test and post-test using the intraclass correlation coefficient (ICC). A two-way mixed model with consistency specification was applied to analyze the data. The findings indicated a high degree of reliability for the measurements at both testing occasions. The average measure ICC for the pre-test was calculated as .996, with a 95% confidence interval between .993 and .997 ($F(5.59) = 224.892$, $p < .001$). These results are summarized in Table 4.

Table 4. Intraclass Correlation Coefficient for pre-test

	Intraclass Correlation [b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.991[a]	.985	.995	224.892	69	69	.000
Average Measures	.996[c]	.993	.997	224.892	69	69	.000

A two-way mixed-effects model (random person effects, fixed measure effects) was used. The estimator is unaffected by the presence of the interaction effect. Type C ICC based on consistency definition was computed, excluding between-measure variance from the denominator.

For the post-test, the average measure ICC was .991 (95% CI [.984, .994], $F(69,69) = 105.642$, $p < .001$), as shown in Table 5.

Table 5. Intraclass Correlation Coefficient for post-test

	Intraclass Correlation [b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.981[a]	.969	.989	105.642	69	69	.000
Average Measures	.991[c]	.984	.994	105.642	69	69	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Taken together, the results demonstrate a high level of consistency between the two raters across both the pre-test and post-test assessments.

3. Group Homogeneity Checks

Prior to the main analyses, the homogeneity of the three groups was examined to ensure their comparability at baseline. Specifically, three one-way analyses of variance (ANOVAs) were performed to compare the groups on (a) their OOPT scores as a measure of general language proficiency, and (b) their pre-test writing scores as assigned by Rater 1 and Rater 2. The outcomes of these homogeneity checks are summarized in Table 6.

Table 6. Homogeneity of Groups in Terms of Language Proficiency and Pre-Test Scores

		N	Mean	SD	Std. Error	df	Mean Square	F	Sig.
OOPT	Instructor-Assessment Group	20	87.0909	5.57980	1.18962	2	143.498	5.470	.007
	Peer-Assessment Group	25	89.0526	5.23316	1.20057	67			
	Self-Assessment Group	25	92.3684	4.39963	1.00934	69			
	Total	70	89.3833	5.49604	.70954				
Pre-test Rater 1	Instructor-Assessment Group	20	67.0000	4.15188	.88518	2	567.130	32.087	.000
	Peer-Assessment Group	25	72.4737	5.43004	1.24574	57			
	Self-Assessment Group	25	77.5263	2.52473	.57921	69			
	Total	70	72.0667	6.02500	.77782				
Pre-test Rater 2	Instructor-Assessment Group	20	67.0909	4.03448	.86015	2	523.537	30.368	.000
	Peer-Assessment Group	25	72.2632	5.48576	1.25852	67			
	Self-Assessment Group	25	77.2105	2.34708	.53846	69			
	Total	70	71.9333	5.86534	.75721				

a. OOPT Proficiency Comparisons

The one-way ANOVA results indicated significant differences in OOPT scores across the three groups, $F(2, 67) = 5.470$, $p = .007$. Post hoc comparisons using Tukey's b test revealed that the Self-Assessment group ($M = 92.36$) had a statistically higher mean score than both the Instructor-Assessment group ($M = 87.09$) and the Peer-Assessment group ($M = 89.05$). However, no significant difference was found between the Instructor-Assessment and Peer-Assessment groups.

b. Pre-test Writing Score Comparisons

With respect to the pre-test writing scores, separate one-way ANOVAs were conducted for the scores assigned by Rater 1 and Rater 2. The results showed statistically significant differences among the three groups for both raters: F Rater 1 (2, 67) = 32.08, p = .000; F Rater 2 (2, 67) = 30.36, p = .000. In both sets of ratings, the Self-Assessment group ($M = 77.21$) significantly outperformed the other two groups. Additionally, the Peer-Assessment group ($M = 72.26$) achieved significantly higher pre-test scores than the Instructor-Assessment group ($M = 67.09$).

4. Inferential Statistics

a. Quantitative Phase Analysis (True-Experimental Phase)

To analyze the data from the true-experimental phase of the study, a mixed-design Analysis of Covariance (ANCOVA) was employed. This model comprised one within-subject factor (writing performance, measured at two time points: pre-test and post-test), one between-subject factor (group membership: self-assessment, peer-assessment, or instructor-assessment), and two covariates (OOPT scores and pre-test writing scores). Before conducting the primary analyses, the underlying statistical assumptions including normality, homogeneity of variances, and sphericity were tested and met.

b. Within-Subjects Effects

The results of the within-subjects effects analysis indicated that the two-way interaction between writing performance and OOPT was not statistically significant, $F (1, 65) = 1.370$, $p = .247$, $\eta^2 = .024$. This suggests that, despite the presence of statistically significant differences in language proficiency levels across the three groups, these proficiency differences do not explain the variability observed in students' post-test writing scores. On the other hand, the interactions for writing performance × group, $F(1, 65) = 8.08$, $p = .001$, $\eta^2 = .227$, and writing performance × pre-test, $F(1, 65) = 120.169$, $p = .000$, $\eta^2 = .686$, reached statistical significance. Detailed results are presented in Table 8.

Table 8. Tests of Within-Subjects Effects

Source		Type III	df	Mean	F	Sig.	Partial Eta
factor1	Sphericity	506.001	1	506.001	103.337	.000	.653
	Greenhouse-	506.001	1.000	506.001	103.337	.000	.653
	Huynh-Feldt	506.001	1.000	506.001	103.337	.000	.653
	Lower-bound	506.001	1.000	506.001	103.337	.000	.653
factor1 * OOPT	Sphericity	6.708	1	6.708	1.370	.247	.024
	Greenhouse-	6.708	1.000	6.708	1.370	.247	.024
	Huynh-Feldt	6.708	1.000	6.708	1.370	.247	.024
	Lower-bound	6.708	1.000	6.708	1.370	.247	.024
factor1 * Pre2	Sphericity	588.422	1	588.422	120.169	.000	.686
	Greenhouse-	588.422	1.000	588.422	120.169	.000	.686
	Huynh-Feldt	588.422	1.000	588.422	120.169	.000	.686
	Lower-bound	588.422	1.000	588.422	120.169	.000	.686
factor1 * Group	Sphericity	79.198	2	39.599	8.087	.001	.227
	Greenhouse-	79.198	2.000	39.599	8.087	.001	.227
	Huynh-Feldt	79.198	2.000	39.599	8.087	.001	.227
	Lower-bound	79.198	2.000	39.599	8.087	.001	.227
Error(factor1)	Sphericity	269.314	55	4.897			
	Greenhouse-	269.314	55.000	4.897			
	Huynh-Feldt	269.314	55.000	4.897			
	Lower-bound	269.314	55.000	4.897			

c. Within-Group Comparisons

The results revealed statistically significant differences between pre-test and post-test writing performance within each of the three groups, indicating substantial improvement over the course of the intervention. Paired-samples t-test analyses yielded the following findings:

- *Instructor-Assessment Group*: A significant increase in writing scores was observed from pre-test (M = 67.00) to post-test (M = 90.50), $t(21) = -19.580$, $p < .001$.
- *Peer-Assessment Group*: Similarly, this group demonstrated a significant improvement from pre-test (M = 72.47) to post-test (M = 91.10), $t(18) = -11.284$, $p < .001$.
- *Self-Assessment Group*: This group also showed a significant gain, with mean scores rising from 77.52 on the pre-test to 94.78 on the post-test, $t(18) = -29.182$, $p < .001$.

Collectively, all three groups performed significantly better on the post-test than on the pre-test. These within-group differences are illustrated in Figure 1.

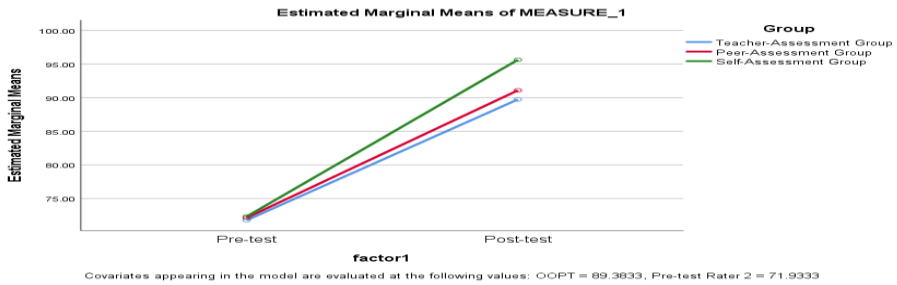


Figure 1. Within-Group Differences in Pre-Test Assessment Compared with Post-Test Assessment of Each Group

With respect to the between-subjects effects, the results revealed significant differences among the three groups across the pre-test and post-test assessments. Furthermore, as each assessment was scored by two independent raters, the multiple comparisons corresponding to each rater are presented separately in Table 9.

d. Between-Subjects Effects

Regarding the between-subjects effects, the findings indicated statistically significant differences among the three groups on both the pre-test and post-test measures. Additionally, given that each assessment was evaluated by two independent raters, the results of the multiple pairwise comparisons for each rater are reported in Table 9.

Table 9. Multiple Comparisons Related to Two Raters on Post-Test Assessment

Dependent Variable		(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Post-test Rater 1	LSD	Instructor-Assessment Group	Peer-Assessment Group	-.60526	.98030	.539	-2.5683	1.3577
			Self-Assessment Group	-4.28947*	.98030	.000	-6.2525	-2.3265
		Peer-Assessment Group	Instructor-Assessment Group	.60526	.98030	.539	-1.3577	2.5683
			Self-Assessment Group	-3.68421*	1.01553	.001	-5.7178	-1.6507
		Self-Assessment Group	Instructor-Assessment Group	4.28947*	.98030	.000	2.3265	6.2525
			Peer-Assessment Group	3.68421*	1.01553	.001	1.6507	5.7178

Post-test Rater 2	LSD	Instructor-Assessment Group	Peer-Assessment Group	-.55981	.93107	.550	-2.4242	1.3046
			Self-Assessment Group	-4.29665*	.93107	.000	-6.1611	-2.4322
		Peer-Assessment Group	Instructor-Assessment Group	.55981	.93107	.550	-1.3046	2.4242
			Self-Assessment Group	-3.73684*	.96453	.000	-5.6683	-1.8054
		Self-Assessment Group	Instructor-Assessment Group	4.29665*	.93107	.000	2.4322	6.1611
			Peer-Assessment Group	3.73684*	.96453	.000	1.8054	5.6683

*. The mean difference is significant at the 0.05 level.

e. Post-Test Comparisons

As shown in the preceding table, the self-assessment group outperformed both the peer-assessment and instructor-assessment groups on the post-test, as measured by both raters. In contrast, no statistically significant differences were found between the instructor-assessment and peer-assessment groups with respect to their post-test writing scores. A boxplot depicting the score distribution for each experimental group is presented in Figure 2.

5. Qualitative Phase: Research Approach and Objectives

The second phase of the study employed a qualitative-exploratory, descriptive, and interpretive framework (Bonilla & Rodríguez, 1997; Burns, 1999; Johnson, 1992; Leedy & Ormrod, 2001), which was distinguished by the inclusion of an intervention. The primary objectives of this phase were to observe, understand, and interpret the role of formative assessment (FA) in influencing students' perceptions of their learning process as well as their attitudes toward this type of assessment. A key interest of the researcher was to examine how FA might shape students' beliefs about their own learning. An additional central concern was to explore students' opinions regarding the implementation of FA and to determine whether it effectively achieved its intended goals.

a. Sources of Student Awareness

The data collected through student interviews alone confirmed the specific situations that gave rise to this awareness. This awareness manifested across a range of contexts, not exclusively those involving formal FA procedures such as assessment tasks, self-assessments, conferences, or remedial activities. It also emerged from classroom

situations unrelated to FA procedures, experiences outside the classroom, and other unspecified circumstances (see Table 10).

Table 10. Awareness of Situations

Situations in which students realized of their learning	Number of students
By FA procedures (self- and peer-assessment)	49
By classroom situations different from FA (self- and peer-assessment)	50
Outside the classroom	49
Unspecific situations	50

Most students also reported experiencing a sense of success through formative assessment, as they became aware of having achieved something, and this awareness emerged during FA procedures. In the interviews, fifty-five students expressed a degree of satisfaction because they realized they had learned a great deal thanks to FA.

Discussion

The first research question sought to determine whether there was any difference in writing performance between the control group (instructor-assessment) and the experimental groups (self-assessment and peer-assessment). The results indicated that the self-assessment group obtained higher post-test scores in both ratings compared to the peer-assessment and instructor-assessment groups. No statistically significant differences were found between the instructor-assessment and peer-assessment groups regarding their post-test writing scores. This finding is consistent with a substantial body of previous research (Birjandi & Hadidi Tamjid, 2012; Birjandi & Siyyari, 2011; Iraji, Enayat, & Momeni, 2016; Liu & Brantmeier, 2019; Mazloomi & Khabiri, 2018; Ratminingsih et al., 2018, among others). Furthermore, this result aligns with a considerable number of studies suggesting that learner involvement in the assessment and feedback process enhances the quality of written drafts (e.g., Jensen & Fischer, 2005; Reiber, 2006; Topping, Smith, Swanson, & Elliot, 2000).

To justify this finding, it can be argued that self-assessment and peer-assessment have enabled writers to integrate the learned standards and criteria for better writing into their own revisions, thereby increasing the quality of their final written drafts. Moreover, participants' engagement in self-assessment and peer-assessment activities is likely to foster their sense of autonomy, agency, and self-regulation, which helped them critically analyze their own drafts and strive to refine their written essays.

The second research question inquired which type of assessment (self-, peer-, or instructor-assessment) would promote the greatest achievement in the writing course. According to the results, all groups demonstrated significant gains from pre-test to post-

test scores. However, the self-assessment group exhibited higher gains compared to the other two groups. The data analysis revealed that the self-assessment group outperformed the other two groups in writing performance as measured by timed-writing essays. This finding concurs with a significant number of previous studies that found self-assessment to be more effective than peer-assessment in enhancing EFL writing performance (Panadero, Brown, & Strijbos, 2016; Bhatti, 2020; Jensen & Fischer, 2005, among others).

Additionally, the results of the study revealed that self-assessment and peer-assessment techniques can serve as effective tools for enhancing learners' intrinsic motivation and improving their self-confidence. This is particularly important in the Iranian EFL context, which heavily depends on the instructor's role in teaching and evaluation. This finding is confirmed by Birjandi and Hadidi Tamjid's (2010) study, in which self-assessment proved to raise students' self-awareness and motivation.

The third research question sought to determine whether the differences in assessment methods (self-, peer-, or instructor-assessment) would be reflected in students' final examination scores. The results demonstrated that both within-group and between-group effects were significant, and different experiments showed significant differences. The analyses indicated statistically significant within-group differences in group performance. Specifically, all groups obtained higher scores on the post-test assessment compared to their pre-test writing performance. This finding is consistent with the results of several studies (e.g., Birjandi & Siyyari, 2011; Liu & Brantmeier, 2019; Delmonte, 1997).

Furthermore, the qualitative phase of the study sought to explore and describe the answers to two questions. These questions aimed to investigate the role of self-assessment implementation and peer-assessment implementation on students' views of their learning. According to the results, nearly all students believed that self-assessment processes, in light of the teacher's guidance, led to their improvement in writing. Moreover, the results revealed that students were better able to recognize their problems and improve their writing when they received feedback from their peers. At some point in instructors' careers, they have failed to recognize the formative nature of assessment, its underlying principles, and the most suitable alternatives for implementing it in the classroom, as suggested by the findings of Arias and Maturana (2005), Bernal and López (2009), and Black and Wiliam (2006). This implementation is not a model but rather an example of the principles that guide formative assessment.

Therefore, formative assessment represents a way to provide opportunities such as those experienced by the participants of this study: they self-assessed their learning in different learning cycles, received teacher feedback in a friendly and democratic

atmosphere on their own initiative, attempted to overcome their problematic areas, and had sufficiently transparent criteria to decide on their readiness for the next learning challenge. The findings of this study are in line with those of Estrada and Vallejo (2006) and Rivers (2001).

In addition, encouraging learners to self-assess their learning first and subsequently promoting their self-regulation were necessary to make learning explicit. The FA proposal somehow conforms with Little's view (1999) that in the autonomous classroom, it is essential that the learner be stimulated to develop an awareness of the aims and processes of learning and to develop a capacity for critical reflection, which also implies having students reflect on their strengths, weaknesses, and progress in various linguistic skills. Consequently, the participants in this study were able to describe language from the perspectives of organizational and pragmatic knowledge and to articulate how they were developing their communicative competence.

CONCLUSION

Formative assessment enables students to develop an understanding of their learning, including the strengths and weaknesses of their communicative competence, and to recognize the contexts—including those that exposed them to FA—in which this awareness was enhanced. Additionally, FA allows them to experience success in their learning. Furthermore, learners perceive formative assessment as a transparent process.

Self-assessment and peer-assessment instruments can be recycled or reused (as instructors commonly do with much of their teaching materials) each time students are assessed. In turn, students learn to evaluate their own learning without having to wait passively for their instructor to grade exams, record scores, and report the results. Later, through brief yet effective instructor-learner dialogues during class time, they can discuss and reach agreement about their learning. This process can become part of regular classroom procedures, like many others, without requiring additional work from the instructor. Making room for formative assessment as part of teachers' instructional practices may be time-consuming initially, but it proves practical and rewarding in the end.

REFERENCES

Akimov, A., & Malin, M. (2020). When old becomes new: A case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education, 45*(8), 1205–1221. https://doi.org/10.1080/02602938.2020.1730301

Allen, L. Q. (2004). Implementing a culture portfolio project within a constructivist paradigm. *Foreign Language Annals, 37*(2), 232–239.

Anasse, K., & Rhandy, R. (2021). Teachers' attitudes towards online writing assessment during Covid-19 pandemic. *International Journal of Linguistics, Literature and Translation, 4*(8), 65–70.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12–19. https://doi.org/10.1080/00405840802577544

Arias, C., & Maturana, L. (2005). Evaluation in foreign languages: Practice and discourse. *Íkala, Revista de Lenguaje y Cultura, 10*(16), 63–91.

Arias, C., Areiza, H., Estrada, L., Marín, J., & Restrepo, E. (2010). *Foreign around the world: Cross-cultural perspectives* (pp. 227–245). University of Hawaii Press.

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education* (10th ed.). Cengage Learning.

Ary, D., Jacobs, L. C., Sorensen, C., & Walker, D. A. (2014). *Introduction to research in education* (9th ed.). Cengage Learning.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*(5), 536–553. https://doi.org/10.1002/sce.1022

Bhatti, A. (2020). Impact of learner-learner rapport on L2 learning: A study of public sector universities in Sindh, Pakistan. *The Asian EFL Journal, 27*(4.6), 204–226.

Birjandi, P., & Hadidi Tamjid, N. (2010). The role of self-assessment in promoting Iranian EFL learners' motivation. *English Language Teaching, 3*(3), 211–220.

Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education, 37*(5), 513–533. https://doi.org/10.1080/02602938.2010.549204

Birjandi, P., & Siyyari, M. (2011). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *Iranian Journal of Applied Linguistics, 13*(1), 23–45.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Bonilla, E., & Rodríguez, P. (1997). *Beyond methods: Research in social sciences*. Norma.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning, 34*(4), 21–42.

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge University Press.

Delmonte, R. (1997). Learning language with a "S.L.I.M." automatic tutor. *Asiatica Venetiana, 2*, 31–52. http://project.cmg.unive.it/slimpage/slimasicc.pdf

Dwiyanti, K. E., & Suwastini, N. K. A. (2021). Assessment for writing skills in online learning. *Lingua Scientia, 28*(1), 8–19.

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning, 32*(5–6), 527–555. https://doi.org/10.1080/09588221.2018.1527361

Estrada, L., & Vallejo, M. (2006). Aiming for a successful summative evaluation in the foreign language class through an effective formative evaluation. *Synergies Venezuela: Revue de didactologie des langues-cultures, 2*, 258–265.

Fahimi, H., Rahimi, A., & Shams, O. (2013). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12–19.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*(2), 175–187. https://doi.org/10.1080/1355800950320212

Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum, R. J., & Köller, O. (2020). Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study. *Assessing Writing, 43*, Article 100420. https://doi.org/10.1016/j.asw.2019.100420

Gipps, C. (1994). Developments in educational assessment: What makes a good test? *Assessment in Education: Principles, Policy & Practice, 1*(3), 283–292.

Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *The Curriculum Journal, 16*(2), 207–223. https://doi.org/10.1080/09585170500136093

Hashemi, M. R., Khodadadi, E., & Yazdanmehr, E. (2009). Learners' evaluation of EFL writing tasks in Iran's ESOL's exam preparation courses. *Journal of English Language Teaching and Learning*.

Herrington, A. J., & Cadman, D. (1991). Peer review and revising in an anthropology course: Lessons for learning. *College Composition and Communication, 42*(2), 184–199.

Inoue, A. B. (2015). *Antiracist writing assessment ecologies: Teaching and assessing writing for a socially just future*. Parlor Press.

Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self- and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies, 6*(4), 716–722.

Jensen, W., & Fischer, B. (2005). Teaching technical writing through student peer-evaluation. *Journal of Technical Writing and Communication, 35*(1), 95–100.

Johnson, D. M. (1992). *Approaches to research in second language learning*. Longman.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction, 20*(6), Article 33. https://doi.org/10.1145/2505057

Lee, I., & Mak, P. (2018). Metacognition and metacognitive instruction in second language writing classrooms. *TESOL Quarterly, 52*(4), 1085–1097. https://doi.org/10.1002/tesq.436

Leedy, P. D., & Ormrod, J. E. (2001). *Practical research: Planning and design* (7th ed.). Merrill Prentice Hall.

Little, D. (1999). Learner autonomy is more than a Western cultural construct. In S. Cotterall & D. Crabbe (Eds.), *Learner autonomy in language learning: Defining the field and effecting change* (pp. 11–18). Peter Lang.

Littlewood, W. (1999). Defining and developing autonomy in East Asian contexts. *Applied Linguistics, 20*(1), 71–94. https://doi.org/10.1093/applin/20.1.71

Liu, H., & Brantmeier, C. (2019). "I know English": Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. *System, 80*, 60–72. https://doi.org/10.1016/j.system.2018.10.013

Livingston, S. A. (2018). *Test reliability – Basic concepts* (Research Memorandum No. RM-18-01). Educational Testing Service.

López, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and training in language assessment. *PROFILE Issues in Teachers' Professional Development, 11*(2), 55–70.

Mazloomi, S., & Khabiri, M. (2018). The impact of self-assessment on language learners' writing skill. *Innovations in Education and Teaching International, 55*(1), 91–100.

McDonald, R., & Savin-Baden, M. (2004). *A briefing on assessment in PBL* (LTSN Generic Centre Assessment Series No. 13).

Meeus, W., Van Petegem, P., & Engels, N. (2009). Validity and reliability of portfolio assessment in pre-service teacher education. *Assessment & Evaluation in Higher Education, 34*(4), 401–413. https://doi.org/10.1080/02602930802062659

Moore, C., O'Neill, P., & Huot, B. (2009). Creating a culture of assessment in writing programs and beyond. *College Composition and Communication, 61*(1), 86–107.

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education, 14*(2), 129–135. https://doi.org/10.1016/j.iheduc.2010.10.001

Mosquera, L. H., & Macías, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal, 17*(2), 302–312.

Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing: A case study from geography. *Innovations in Education and Training International, 32*(4), 324–335.

Nolen, S., McReynolds, L. S., DeComo, R. E., John, R., Keating, J. M., & Wasserman, G. A. (2008). Lifetime suicide attempts in juvenile assessment center youth. *Archives of Suicide Research, 12*(2), 111–123.

Orsmond, P. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International, 41*(3), 273–290.

Page, E. B. (1966). The imminence of grading essays by computer. *The Phi Delta Kappan, 47*(5), 238–243.

Panadero, E., Brown, G. T. L., & 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

Panadero, E., Fernández Ruiz, J., & Sánchez-Iglesias, I. (2026). Formative assessment in the age of generative AI: A systematic review of self- and peer-assessment practices in online writing instruction. *Assessment & Evaluation in Higher Education, 51*(2), 1–24.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Porte, G., & Richards, K. (2012). Focus article: Replication in second language writing research. *Journal of Second Language Writing, 21*(3), 284–293. https://doi.org/10.1016/j.jslw.2012.05.002

Qu, W., & Zhang, C. (2013). The analysis of summative assessment and formative assessment and their roles in college English assessment system. *Journal of Language Teaching and Research, 4*(2), 335.

Ratminingsih, N. M., Marhaeni, A. A. I. N., & Vigayanti, L. P. D. (2018). Self-assessment: The effect on students' independence and writing competence. *International Journal of Instruction, 11*(3), 277–290. https://doi.org/10.12973/iji.2018.11320a

Rivers, W. P. (2001). Autonomy at all costs: An ethnography of metacognitive self-assessment and self-management among experienced language learners. *The Modern Language Journal, 85*(2), 279–290.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation, 11*(1), Article 10. https://doi.org/10.7275/9wph-vv65

Samper, M. V. (2020). *The use of formative assessment in foreign language teaching for motivation and L2 perception*. University of Barcelona.

Spiller, D. (2012). *Assessment matters: Self-assessment and peer assessment*. University of Waikato.

Tartavulea, C. V., Albu, C. N., Albu, N., Dieaconescu, R. I., & Petre, S. (2020). Online teaching practices and the effectiveness of the educational process in the wake of the COVID-19 pandemic. *Amfiteatru Economic, 22*(55), 920–936. https://doi.org/10.24818/EA/2020/55/920

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education, 25*(2), 149–169. https://doi.org/10.1080/713611428

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment, 6*(1), 41–63.

Urbach, D. (2014). Examining the factor structure of writing assessment based on sets of analytical marking criteria. *Procedia – Social and Behavioral Sciences, 141*, 1106–1111.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046

Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology, 32*(3), 164–166.

Zou, D., Wang, M., Xie, H., Cheng, G., Wang, F. L., & Lee, L. K. (2021). A comparative study on linguistic theories for modeling EFL learners: Facilitating personalized vocabulary learning via task recommendations. *Interactive Learning Environments, 29*(2), 270–282.