Evaluating the Influence of Formative and Summative Assessments on EFL Learners' Listening and Reading Comprehension

Mohammad Bahrami (iD)

Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: bahrami_mhm@yahoo.com

Reihaneh Shoghi (iD) (Corresponding Author)

Department of English Language and Literature, Isfahan University, Isfahan, Iran. E-mail: reihaneshoghi@gmail.com

Amir Mahshanian (iD)

Department of English Language and Literature, Isfahan University, Isfahan, Iran. E-mail: mshn_amir@yahoo.com

Abstract

This study investigates the effects of summative, formative, and combined assessments on the development of listening and reading comprehension skills on advanced Iranian EFL learners. Data were collected from 40 advanced EFL learners, aged 19 to 35, from intact classes at a higher education institute, divided into three experimental groups, each receiving one of the assessment types (summative, formative, or a combination of both), along with a control group. Learners' progress in listening and reading comprehension was measured over a four-month period using pre- and post-tests administered at the end of the term. Data analysis involved ANOVA, ANCOVA, and post hoc LSD tests. The findings of this longitudinal study revealed that while the combined application of formative and summative assessments had a significant positive impact on listening comprehension, neither assessment type alone contributed to improvement in this area. In contrast, the results showed that all forms of assessment—whether summative, formative, or combined—led to notable improvements in reading comprehension. These results highlight the differential effects of assessment types on listening and reading skills, suggesting implications for assessment practices in EFL contexts.

Introduction

Assessment is broadly understood as the process of making judgments about learners' performance and monitoring their progress (Taras, 2005). This definition, though straightforward, draws attention to the two primary forms of assessment—formative and summative—both of which have sparked extensive debate regarding their effectiveness (Black et al., 2010). One of the most widely recognized definitions of formative assessment comes from Black and Wiliam (1998), who describe it as "activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities" (Black & Wiliam, 1998, 2). In contrast, summative assessment is generally characterized as a form of evaluation that is "almost always graded, typically less frequent, and occurs at the conclusion of instructional segments" (Dixson & Worrell, 2016, 6).

The literature on formative and summative assessment highlights their overlap and advocates for integration (Pacheco, 2022). Pacheco (2022) also emphasized that in language learning, formative and summative assessments work together to provide timely feedback and measure learners' achievements across various criteria.

In the last two decades, particularly after Black and Wiliam's study in (1998) on positive aspects of formative assessment, extensive research has explored the effects of both formative and summative assessments in education (e.g., Covic & Jones, 2008; Ghiatau, 2011). Some studies advocate for the advantages of formative assessment (e.g., Covic & Jones, 2008), while others argue that summative assessment yields more reliable results, claiming that formative assessment offers "an incomplete vision" (Dixson & Worrell, 2016).

By the early 20th century, the importance of formative assessment in English as a Second Language (ESL) contexts gained attention (Gan & Leung, 2020). However, studies like Gattullo (2000) focused mainly on children, overlooking adult learning. Additionally, research by Harris (2007), Hanves (2012), and Ghiatău (2011) questioned the validity of their findings due to reliance on student questionnaires. In contrast, other researchers highlighted the significance and reliability of summative assessment over formative assessment (e.g., Lau, 2016; Nieminen et al., 2021; Trotter, 2006). Contrary to numerous studies that explored the effects of various types of assessment on learning in general, only a limited number have specifically examined the impact of different assessment types on the development of listening comprehension (e.g., Buck, 2001; Campione, 1989; Ghahremani, 2013).

While research has extensively explored the effects of different assessment types on listening comprehension, fewer studies have focused on their impact on reading comprehension (e.g., Chen, 2016; Sönmez & Cetinkaya, 2022). Notably, Sönmez and Cetinkaya found that formative assessment strategies significantly enhance reading comprehension outcomes. Although the effects of formative and summative assessments on language learning in educational and ESL contexts have been documented (e.g., Gattullo, 2000; Jiang, 2014), their influence on listening and reading comprehension remains underexplored. This study aims to

address this gap by examining how both assessment types affect EFL learners' listening and reading comprehension.

By focusing specifically on these aspects, this research addresses a critical gap in the current literature, contributing to a deeper understanding of the ways in which assessment practices can be optimized to enhance comprehension skills among EFL learners. This study's findings are expected to inform educators and policymakers about effective assessment strategies that can better support language acquisition and comprehension in EFL contexts.

This present study is extensively grounded in constructivist learning theory, positing that learners build understanding through experiences and reflection (Vygotsky, 1978). The assessment types examined align with this theory by offering ongoing feedback and opportunities for reflection. Specifically, formative and summative assessments are viewed as means to enhance teaching and learning, applying the Assessment for Learning approach that emphasizes using assessment to promote learning rather than just measure it (Black & Wiliam, 1998; Wiliam, 2011).

1. Literature Review

1.1. Formative Assessment and Receptive Language Skills

Formative assessment is crucial for enhancing EFL learners' listening comprehension. Research shows that self-regulatory practices lead to better listening skills as Ashraf et al. (2013) highlight the value of active learning environments that engage learners in tasks beyond simple listening. Additionally, formative assessments can be customized to address various learning styles. Techniques like partial dictation have proven effective in improving listening comprehension by helping learners connect phonemic sounds (Abdolmajid et al., 2021).

Formative assessments enhance listening comprehension by fostering metacognitive strategies, enabling learners to better manage their listening processes (Teo et al., 2022). Research indicates that strategies like planning and self-evaluation significantly improve listening outcomes (Safa & Motaghi, 2021). Additionally, using multimedia resources in assessments caters to diverse learner preferences and boosts listening skills (Kabri, 2023).

Wijaya (2022) emphasizes that formative assessments help EFL teachers identify students' strengths and weaknesses, creating a positive learning environment. This is supported by Huseynova (2023), who found that both teachers and students recognize the value of formative assessments for improving reading skills. Yazidi (2023) emphasizes the importance of effective feedback in enhancing reading comprehension.

1.2. Summative Assessment and Receptive Language Skills

Summative assessments significantly impact learners' anxiety and self-efficacy in listening comprehension. Bayat's (2017) study highlights that formative assessment reduces anxiety and improves listening efficacy in EFL learners, indicating that the type of assessment influences emotional responses and performance. Similarly, Alshahrani (2023) noted that extensive listening activities can affect listening anxiety based on their execution.

Additionally, summative assessments are vital for evaluating reading comprehension in EFL learners. Hosseini and Ghabanchi (2014) demonstrated that portfolio assessments, a type of summative assessment, enhance both reading comprehension and motivation.

Moreover, various studies highlight the significant role of summative assessments alongside formative assessments. Ismail et al. (2022) found that summative assessments impact academic motivation and self-regulation in EFL learners, potentially inducing anxiety that hampers reading comprehension performance. Mahshanian et al. (2019) also noted that summative assessments offer valuable feedback for identifying strengths and weaknesses in reading skills. Additionally, Zandi (2022) pointed out that an overemphasis on summative assessments by EFL instructors may limit understanding of students' abilities and reduce their engagement with reading materials.

In addition to the psychological aspects, the pedagogical implications of summative assessments are significant. The study by Pangestu et al. (2021) emphasizes that reading comprehension is a critical skill for EFL learners, and effective assessment strategies, including summative assessments, are essential for fostering this skill through contextualizing assessment items and using various question formats for a more thorough evaluation of their abilities.

1.3. The Integrated Approach

Mahshanian et al. (2019) argue that an integrated approach to formative and summative assessments can result in higher achievement levels in reading comprehension than either assessment type alone. The continuous feedback provided through these types of assessments allows learners to better understand their strengths and weaknesses, fostering a more personalized learning experience that is crucial for developing reading skills (Ismail et al., 2022).

Moreover, formative and summative assessments are believed to promote metacognitive strategies which can lead to improved comprehension outcomes for EFL learners. Ahmadi et al. (2013) emphasize that metacognitive reading strategy awareness is essential for enhancing reading comprehension among EFL learners.

The effectiveness of assessments is influenced by prior instructional methods (Yildirim et al., 2024). Al-Qahtani (2015) found that explicit teaching of discourse markers improves reading comprehension, highlighting the importance of quality instruction. Thus, while assessments offer insights into comprehension abilities, their effectiveness depends on the instructional strategies used, and as Zandi (2022) suggested, a broader range of assessment types should be employed to reflect diverse classroom practices. To guide the investigation, the following research questions were formulated:

1. What is the effect of different assessment types on the development of listening comprehension among EFL learners?

2. How does the implementation of individual assessment types, as well as their interaction, significantly affect EFL learners' reading comprehension performance?

By addressing these questions, the study aims to contribute to the understanding of effective assessment practices in EFL contexts, particularly concerning their influence on advanced learners' comprehension skills.

2. Method

2.1. Design

Employing a quasi-experimental design, this study was conducted over a four-month period within an English as a Foreign Language (EFL) context in Iran, focusing on advanced-level learners. The primary objective was to examine the effects of two types of assessment—formative and summative—and their integration on the improvement of these learners' listening and reading comprehension skills. The study included a control group that did not receive the same assessment interventions. The independent variables were the types of assessment (formative and summative assessments, both separately and in combination), while the dependent variables were the students' listening and reading comprehension skills. The instructional intervention consisted of 100 sessions, each lasting 90 minutes, during which various assessment strategies were implemented to gauge their impact on student performance.

2.2. Participants

The study involved 40 advanced EFL learners, aged 19 to 35, from intact classes at a higher education institute. Four groups of 10 participants each were formed based on a pre-assessment of their listening and reading comprehension skills, exhibiting comparable levels of proficiency in these areas.

Three groups received tailored instructional treatments: one focused on formative assessment, another on summative assessment, and the third on a combination of both. The fourth group acted as a control, receiving no specific assessment treatment. Detailed background information about the participants and instructors is available in the following tables:

Table 1. *Demographic and Background Characteristics of EFL Learner Participants*

Classes	Groups	Number of members	Male	Female	Age	Proficiency Level
1	Summative	10	5	5	19-35	Advanced
2	Formative	10	4	6	19-35	Advanced
3	Interaction	10	6	4	19-35	Advanced
4	Control group	10	5	5	19-35	Advanced

Table 2. *Demographic and Professional Profiles of Instructor Participants*

Instructors	Age	Education	Major	Experience (years)	Gender
Ins 1	26	MA	TEFL	8	M
Ins 2	32	MA	TEFL	10	F
Ins 3	40	MA	TEFL	15	M
Ins 4	29	MA	TEFL	6	F

2.3. Materials and Instruments

2.3.1. Instructional Materials

The primary instructional materials utilized for this study were Passages Series 1 and 2 (Richards, 2014), comprising twelve units aimed at enhancing various communicative skills, including grammar, vocabulary, speaking, listening, writing, and reading. This two-level multi-skill course targets rapid advancement for high-intermediate to advanced EFL learners.

These materials were selected based on expert evaluations, confirming their alignment with the learners' previous coursework, particularly with *Interchange* Series, the textbooks used prior to the study's pre-test phase.

Each unit of the *Passages* series supports the course objectives, commencing with a "Starting Point" section that introduces the new grammar in context. This is followed by vocabulary and speaking sections, contextualized grammar practice, and interactive speaking activities. Listening sections focus on enhancing comprehension skills, while reading and writing sections develop academic writing and critical thinking through authentic texts.

2.3.2. Testing Materials

The evaluation process of learners' performance was accomplished through using tests and quizzes from the *Passages Series*, designed to assess various communicative competencies and vetted by experts. An achievement test by Richards and Sandy (2014) served as both the pre-test and post-test to compare learners' abilities before and after the intervention.

Participants took the pre-test before treatment sessions and the post-test afterward, assessing listening, reading comprehension, and core language skills. Additionally, 12 quizzes aligned with formative assessment principles were administered after every two units to provide ongoing feedback for continuous improvement. Details on assessment items and frequencies are in Appendix B.

Test item validation involved expert consultation in EFL assessment to ensure appropriateness for the study's intact classes. The reliability of the pre-test and quizzes was calculated using the Kuder-Richardson Formula 21 (KR-21) to confirm consistency, with resulting reliability indices provided below.

Table 3. *Reliability of the Utilized Tests*

Test	Reliability
Mid-term exam	76.4%
Quizzes	65%
Grammar items	78.6%
Vocabulary items	73.5%
Listening items	70.1%
Reading items	72%

2.4. Procedure

2.4.1 Participants' Homogeneity

Before the commencement of the study, all participants were administered a pretest from the Passages Series. Approximately 10% of the participants were excluded for high pretest scores.

The same test was readministered at the end of the course to evaluate the impact of different assessment types on listening and reading comprehension.

Instructors participated in briefing sessions on assessment requirements prior to the initiation of the instructional sessions, conducted three times a week for 90 minutes. Four intact groups were used: one received only summative assessment, one only formative assessment, one experienced both types, and one served as a control group.

2.4.2. Formative Assessment Group

In the formative assessment group, instructors focused on providing feedback and encouraging self-evaluation and reflection for continuous development, following the Assessment Reform Group's principles (2002). Notably, instead of final tests, they conducted quizzes after every two units and did not give verbal or non-verbal feedback or make performance comparisons. They allowed flexibility regarding mistakes, with only a pretest and posttest administered.

2.4.3. Summative Assessment Group

Conversely, the summative assessment group evaluated learners through midterm and final exams, reporting scores to the institution's director. Feedback was limited to binary responses, with no immediate evaluations during instruction. Unlike the formative group, learners did not participate in peer evaluations and received no detailed feedback. They took midterm exams after six units and a comprehensive final exam after Unit 12.

2.4.4. Interaction Assessment Group

The interaction group underwent the same midterm and final assessments as the summative group, supplemented by quizzes and the feedback mechanisms employed in the formative group. In essence, participants in the interaction group benefited from a combination of formative assessment strategies and traditional summative evaluations, thereby enriching their learning experience.

2.4.5. Control Group

The control group instructors conducted their classes without prior training in various assessment types. It is noteworthy that the instructor assigned to this group did not participate in the initial briefing sessions.

All instructional sessions were recorded, and selected audio recordings were transcribed to ensure adherence to study protocols. Instructors were also required to report their assessment procedures to the researchers to maintain consistency and integrity throughout the study.

2.5. Data Analysis

An independent samples t-test was conducted to assess the pretest performance of learners across the control group (CG) and the formative (FG), summative (SG), and interaction groups (IG). The results showed no significant differences in pretest scores among the groups, indicating that all learners had similar levels of proficiency before the intervention.

Subsequent sections of this paper will detail the improvements in listening and reading comprehension attributed to the three assessment methods employed: summative, formative, and a combination of both.

3. Results

3.1. Improvement in Listening Comprehension
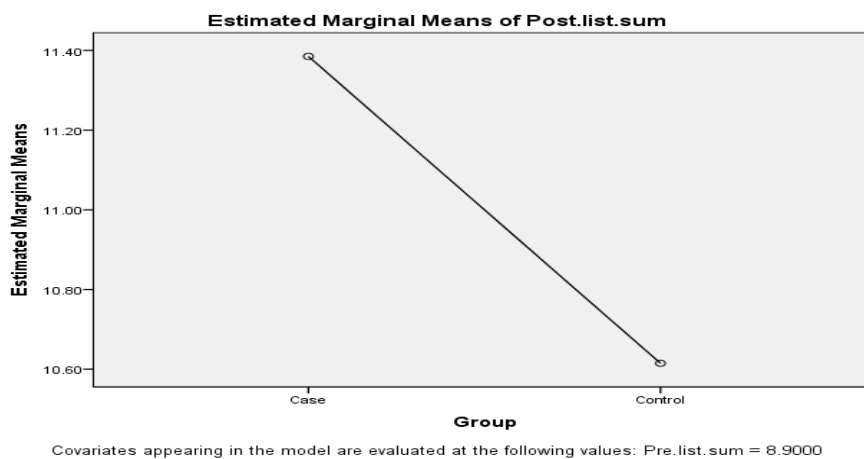
3.1.1. Summative Group (SG)

Table 4 displays the descriptive statistics for learners' scores on the listening comprehension test. The data indicate that the mean score for the summative group (SG) is significantly higher ($M = 11.10$) compared to the control group (CG), which recorded a lower mean score ($M = 10.90$).

Table 4. *Descriptive Statistics of Summative Group Listening Post-Test*

Group	Mean	Std. Deviation	N
Case	11.1000	.73786	10
Control	10.9000	1.28668	10
Total	11.0000	1.02598	20

Efforts were made to control extraneous variables, but ANCOVA results exhibited that high and low performers on the listening comprehension pretest remained categorized as such even after the treatment ($p = 0.005$). This trend was consistent across groups. Further ANCOVA analysis for the summative group (SG) is detailed in Appendix C.

The ANCOVA results (F (1, 17) = 3.338, $p = 0.085$) revealed no significant difference in post-test performance between the summative group (SG) and the control group (CG). While the SG had a slightly higher mean score ($M = 11.10$) than the CG ($M = 10.90$), this difference was not statistically significant. Thus, the summative assessment did not improve listening comprehension scores, as illustrated in Figure 1.

Figure 1. *Lack of significant improvement in listening comprehension within the Summative Assessment Group (SG)*
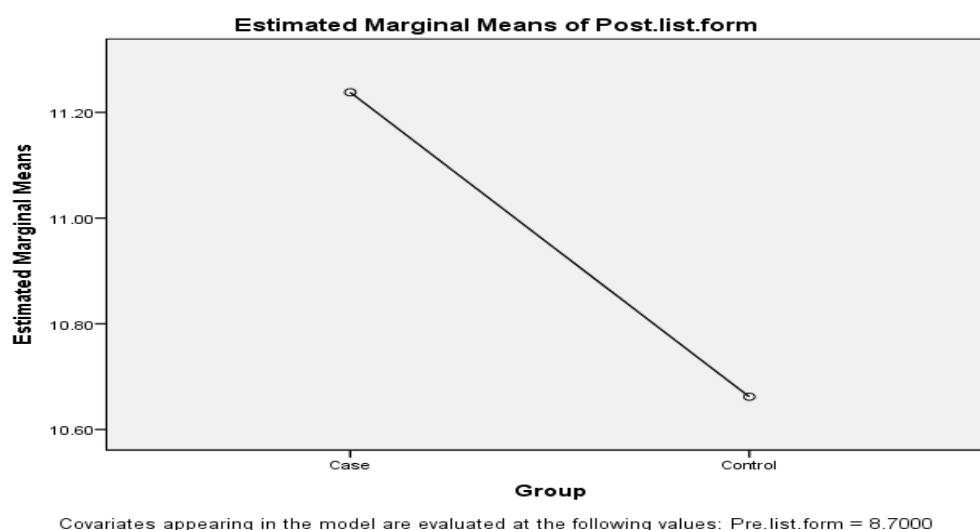
3.1.2. Formative Group (FG)

Table 5 presents the descriptive statistics for the learners' scores in the listening comprehension test within the formative group (FG). As illustrated in the table, the mean score for the FG ($M = 11.00$) exceeds that of the control group (CG) ($M = 10.90$).

Table 5. *Descriptive Statistics of Formative Group Listening Post-Test*

Group	Mean	Std. Deviation	N
Case	11.0000	0.94281	10
Control	10.9000	1.28668	10
Total	10.9500	1.09904	20

To assess the effectiveness of the treatment in the formative group (FG), an analysis of covariance (ANCOVA) was conducted, with results presented in Appendix D.

The ANCOVA results (F (1, 17) = 1.271, *p* = 0.275) demonstrate no significant difference in listening comprehension test performance between the formative group (FG) and the control group (CG). Although FG had a slightly higher mean score (*M* = 11.00) compared to CG (*M* = 10.90), the formative assessment did not lead to meaningful improvements. Figure 2 reflects this lack of significant enhancement in the FG.

Figure 2. *Lack of significant improvement in listening comprehension within the Formative Assessment Group (FG)*

3.1.3. Interaction Group (IG)

Table 6 presents the descriptive statistics for the learners' scores on the listening comprehension test within the interaction group (IG). Notably, the mean score in IG (*M* = 11.80) is higher than that of the control group (CG), which had a mean score of (*M* = 10.90).

Table 6. *Descriptive Statistics of Interaction Group Listening Post-Test*

Group	Mean	Std. Deviation	N
Case	11.8000	.91894	10
Control	10.9000	1.28668	10
Total	11.3500	1.18210	20

To evaluate the treatment's effectiveness in the interaction group (IG), an ANCOVA was performed, with results in Appendix E.

The ANCOVA results indicate a significant relationship between IG and the control group (CG) in listening comprehension test performance. This suggests that combining formative and summative assessments led to notable improvements in scores, while using either method alone

did not produce similar results. Figure 3 highlights the significant gains in listening comprehension for the interaction group.
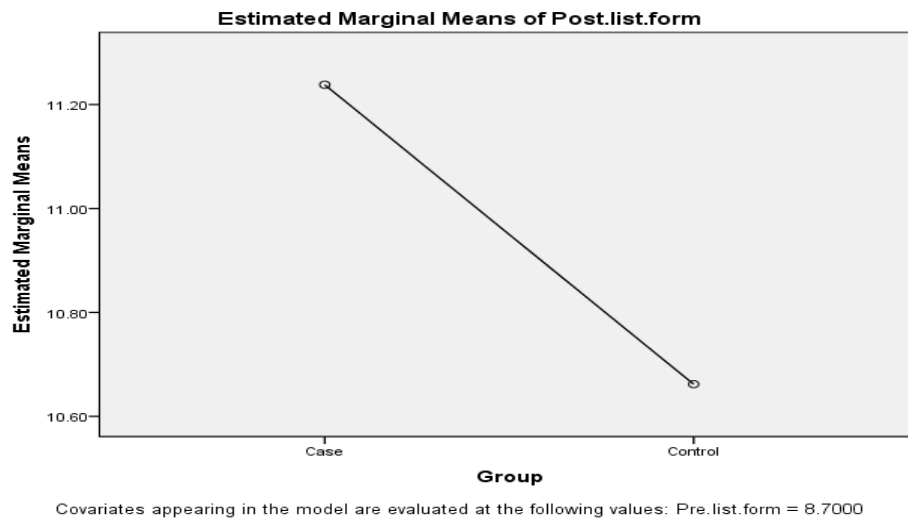
Figure 3. *Improvements in listening comprehension in the Interaction Group (IG)*

3.1.4. Summary of Improvement in Listening Comprehension in Three Groups (SG, FG, IG).

The mean and *SD* were computed for all groups and are summarized in table 7 below.

Table 7. *Descriptive Statistics of Three Groups Post-Test Listening Comprehension*

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval		Score Range	
					Lower Bound	Upper Bound		
FG	10	11.00	.94281	.29814	10.3256	11.6744	Min	Max
SG	10	11.10	.73786	.23333	10.5722	11.6278	10.0	12.0
IG	10	11.80	.91894	.29059	11.1426	12.4574	10.0	12.0
Total	30	11.30	.91539	.16713	10.9582	11.6418	10.0	13.0

A post-Hoc LSD test was conducted to compare listening comprehension scores among the three groups (SG, FG, and IG). The results yielded no significant differences in mean scores ($p > .05$). Since FG and SG had nearly identical scores and neither outperformed the control group (CG), it can be concluded that neither assessment method improved listening comprehension. Results are shown in Table 8.

Table 8. *Post-Hoc-LSD test for Comparing Three Groups in Listening Comprehension Improvement*

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
FG	SG	-.10000	.38968	.799	-.8996	.6996
	IG	-.80000*	.38968	.050	-1.5996	-.0004
SG	FG	.10000	.38968	.799	-.6996	.8996
	IG	-.70000	.38968	.084	-1.4996	.0996
IG	FG	.80000*	.38968	.050	.0004	1.5996
	SG	.70000	.38968	.084	-.0996	1.4996

The post-Hoc LSD test results indicated that learners in both the Summative Group (SG) and Formative Group (FG) did not experience significant changes in their listening comprehension scores after treatment. Although the Interaction Group (IG) performed better than the Control Group (CG), their improvement was not significantly different from that of the SG or FG. An ANOVA was conducted to highlight the relationship between the IG and FG/SG, with results shown in Table 9 below.

Table 9. *ANOVA for Listening Improvement in Three Groups*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.800	2	1.900	2.502	.101
Within Groups	20.500	27	.759		
Total	24.300	29			

The results of ANOVA (Table 9) suggested that there is no significant relationship between the mean scores of the three groups and that of the Control Group (CG). Therefore, the following conclusions can be drawn about improvements in listening comprehension:

1. The Summative Assessment method, on its own, does not lead to improvements in learners' listening comprehension.

2. The Formative Assessment method, on its own, does not lead to improvements in learners' listening comprehension.

3. The Formative Assessment method is as ineffective as the Summative Assessment method concerning learners' improvements in listening comprehension.

4. The combination of both Summative and Formative Assessment methods is the only approach that facilitates improvements in learners' listening comprehension.

To illustrate the effectiveness and ineffectiveness of the three assessment methods for improving listening comprehension based on their post-test means, Figure 4 is presented below.
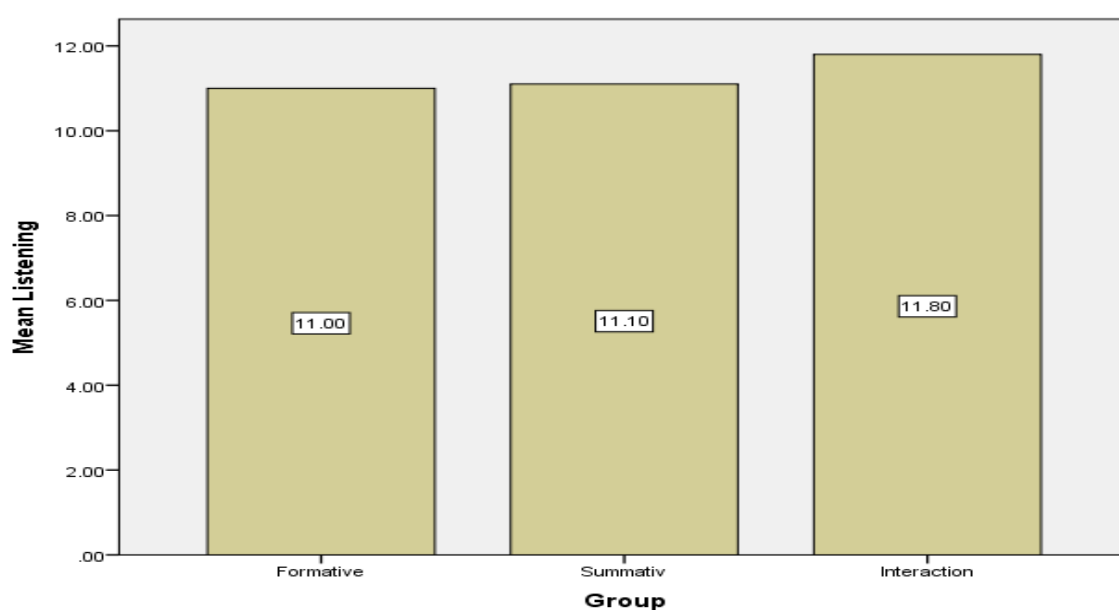
Figure 4. *Mean difference in three groups in listening comprehension improvement*

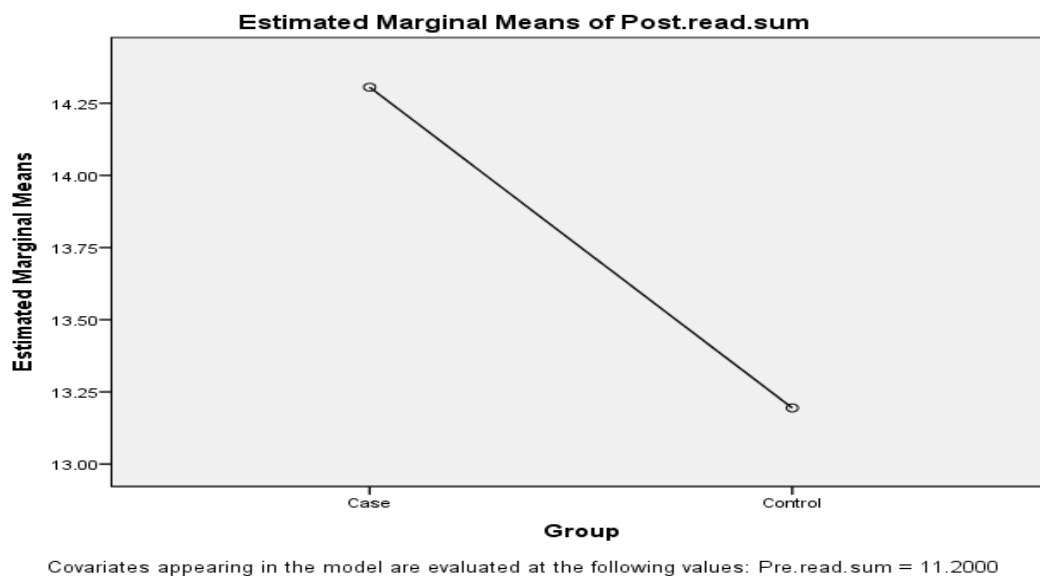### 3.2. Improvement in Reading Comprehension

3.2.1. Summative Group (SG)

Table 10 presents the descriptive statistics for the learners' scores in the reading comprehension test within the Summative Group (SG). The table indicates that the mean score for SG (M = 14.4) is slightly higher than that of the Control Group (CG) (M = 13.1).

Table 10. *Descriptive Statistics of Summative Group Post-Test Reading Comprehension*

Group	Mean	Std. Deviation	N
Case	14.4000	1.26491	10
Control	13.1000	1.72884	10
Total	13.7500	1.61815	20

The results of the ANCOVA conducted to evaluate the effectiveness of the treatment in the Summative Group (SG) are presented in Appendix F.

The ANCOVA results indicate a significant effect of the treatment on reading comprehension scores between the Summative Group (SG) and the Control Group (CG), $F_{(1, 17)}$ = 4.997, p = 0.039. Although the mean score for SG (M = 14.4) was higher than that of CG (M = 13.1), the improvement in reading comprehension was minimal. Figure 5 illustrates the improvement in reading comprehension scores within the Summative Group.

Covariates appearing in the model are evaluated at the following values: Pre.read.sum = 11.2000

Figure 5. *Improvements in reading comprehension in the Summative Group (SG)*

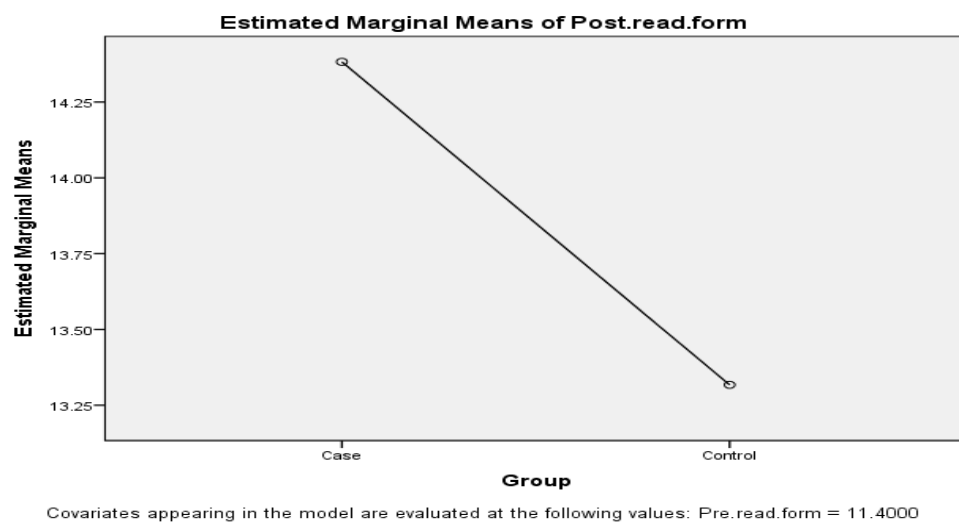### 3.2.2. Formative Group (FG)

Table 11 presents the reading comprehension test scores for the Formative Group (FG) and Control Group (CG). The FG had a mean score of 14.60 (SD = 1.17), slightly higher than the CG's mean of 13.10 (SD = 1.73), indicating that FG learners showed improvement in reading comprehension after the treatment (see Table 11). A similar trend was noted for FG in Table 11.

Table 11. *Descriptive Statistics of Formative Group Post-Test Reading Comprehension*

Group	Mean	Std. Deviation	N
Case	14.6000	1.17379	10
Control	13.1000	1.72884	10
Total	13.8500	1.63111	20

Results of the Analysis of Covariance (ANCOVA) conducted to investigate the effectiveness of the treatment in the Formative Group (FG) are presented in Appendix G.

The ANCOVA results indicate a significant difference between FG and CG in reading comprehension performance, ($F (1,17) = 5.196$, $p = 0.036$). This suggests that while there is a statistically significant improvement in reading comprehension scores in FG compared to CG, the improvement may not be practically significant. Figure 6 illustrates the improvement in reading comprehension scores in FG.

Estimated Marginal Means of Post.read.form

Covariates appearing in the model are evaluated at the following values: Pre.read.form = 11.4000

Figure 6. Improvement in reading comprehension in the Formative Group (FG)
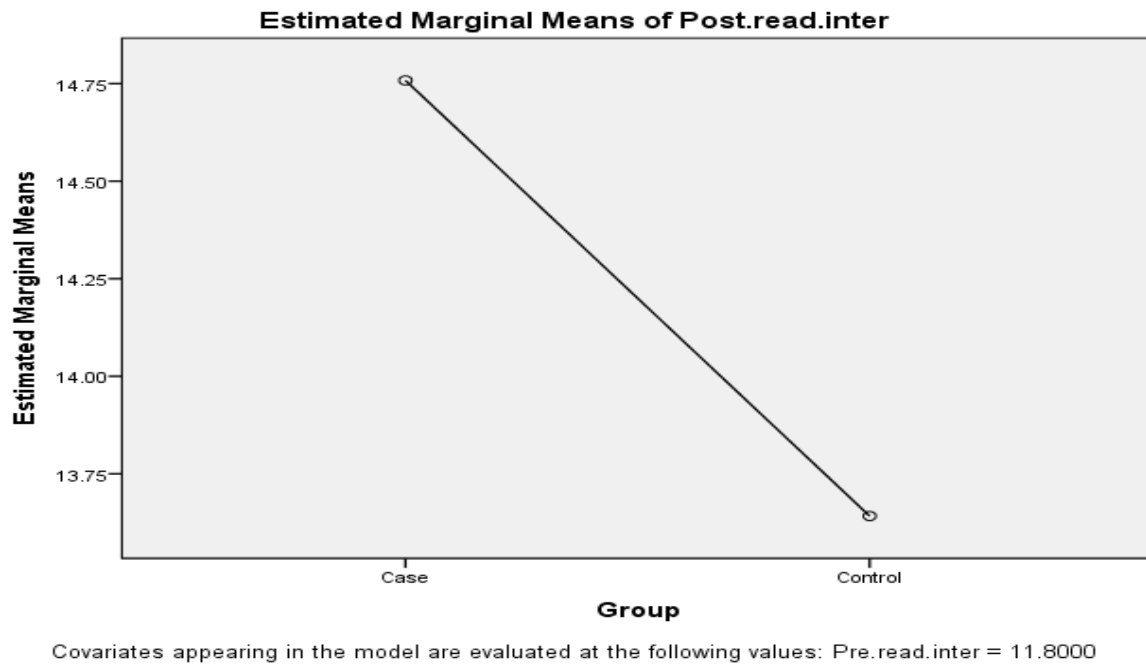
3.2.3. Interaction Group (IG)

Table 12 presents the descriptive statistics for the learners' scores in listening comprehension test. It shows that the mean of scores in IG ($M = 15.3$) is higher than that of CG ($M = 13.1$), and that learners in IG outperformed those in CG which means their reading comprehension improved after the treatment (see table 12). This was also the case for SG, and FG (tables 10 and 11).

Table 12. *Descriptive Statistics of Interaction Group Post-Test Reading Comprehension*

Group	Mean	Std. Deviation	N
Case	15.3000	1.15950	10
Control	13.1000	1.72884	10
Total	14.2000	1.82382	20

To ensure the effectiveness of the treatment in IG, analysis of covariance was run and results are presented in Appendix H.

As indicated by the results, there is a statistically significant relationship between the Interaction Group (IG) and the Control Group (CG) concerning their performance on the reading comprehension test, F (1,17) = 4.663, *p* = .045). However, despite this significant finding, the improvement in reading comprehension scores in IG is not considered substantial, similar to the results observed in the Summative Group (SG) and the Formative Group (FG). Figure 7 illustrates the improvement in reading comprehension scores across the different groups.

Estimated Marginal Means of Post.read.inter

Covariates appearing in the model are evaluated at the following values: Pre.read.inter = 11.8000

Figure 7. *Improvement in reading comprehension in the Interaction Group (IG)*

3.2.4. Summary of Improvement in Reading Comprehension in 3 Groups (SG, FG, IG)

The means and standard deviations (SD) for the reading comprehension post-test scores across the three groups—Summative Group (SG), Formative Group (FG), and Interaction Group (IG)—are presented in Table 13.

Table 13. *Descriptive Statistics of Three Groups Post-Test Reading Comprehension*

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean			
					Lower Bound	Upper Bound		
FG	10	14.600	1.1737	.3711	13.7603	15.4397	13.0	16.0
SG	10	14.400	1.2649	.4000	13.4951	15.3049	13.0	17.0
IG	10	15.300	1.1595	.3666	14.4705	16.1295	14.0	18.0
Total	30	14.766	1.2228	.2232	14.3100	15.2233	13.0	18.0

The post-Hoc LSD test indicated no statistically significant differences between the mean scores of the three groups (*p* > .05), suggesting that the assessment methods (summative, formative, or both) did not notably affect learners' reading comprehension improvement. All groups showed similar outcomes, reflecting a slight enhancement in reading comprehension scores. Detailed results are in Table 14 below.

Table 14. *Post-Hoc-LSD Test for Comparing Three Groups in Reading Comprehension Improvement*

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Formative	Summative	.20000	.53679	.712	-.9014	1.3014
	Interaction	-.70000	.53679	.203	-1.8014	.4014
Summative	Formative	-.20000	.53679	.712	-1.3014	.9014
	Interaction	-.90000	.53679	.105	-2.0014	.2014
Interaction	Formative	.70000	.53679	.203	-.4014	1.8014
	Summative	.90000	.53679	.105	-.2014	2.0014

The results in Table 15 depict learners in the Summative Group (SG), Formative Group (FG), and Interaction Group (IG) all improved their reading comprehension scores due to the assessment methods used. However, while these groups outperformed the control group, they did not significantly differ from each other. An ANOVA confirmed that the differences among the three groups were not significant, as summarized in Table 15.
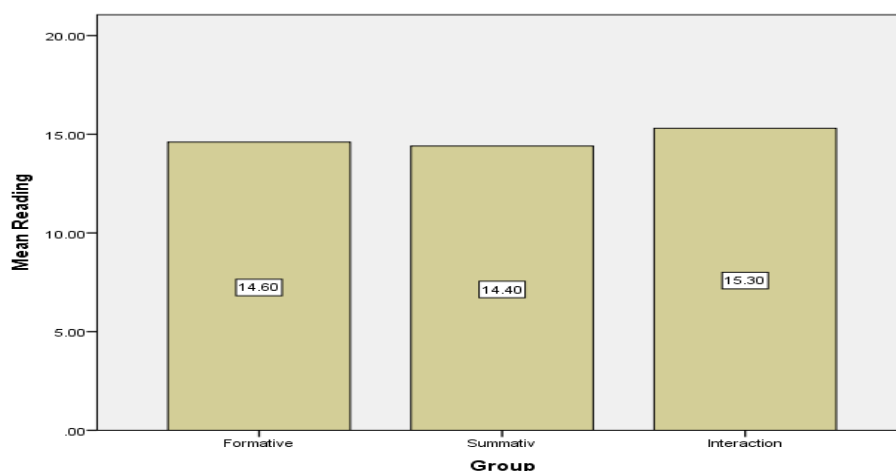
Table 15. *ANOVA for Reading Comprehension Improvement in Three Groups*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.467	2	2.233	1.550	.231
Within Groups	38.900	27	1.441		
Total	43.367	29			

As illustrated in Table 15, there is no significant relationship among the mean scores of the three groups (p = .231). Based on the findings presented in Tables 12 to 15, the following concluding remarks can be made:

1. Assessment, irrespective of its type (i.e., summative, formative, or an interaction of both), is effective in improving learners' reading comprehension scores.

2. Summative, formative, and interaction methods of assessment impact learners' reading comprehension improvement similarly.

To visually depict the effectiveness of all assessment methods in enhancing reading comprehension, Figures 8 and 9 are presented below.

Figure 8. *Mean difference in three groups in reading comprehension improvement*

4. Discussion

The present study aimed to investigate the effects of different assessment methods, addressing gaps identified in prior research within the domain of language education. As previously highlighted, the available literature has explored the relationship between assessment methods and learning outcomes, particularly emphasizing the efficacy of formative assessment.

Several noteworthy strengths of the current study merit discussion. Firstly, it employed quantitative analysis to evaluate both summative and formative assessment methods, contrasting with earlier studies that mainly used qualitative approaches (e.g., Ghiatău et al., 2011; Havnes et al., 2012; Taras, 2008). Secondly, unlike many studies that favor formative assessment, this research did not prioritize any specific method (e.g., Covica & Jones, 2008; Gan & Leung, 2020). Third, the study concentrated on receptive skills, specifically listening and reading comprehension, offering a more detailed analysis compared to earlier research that generalized the effects of various assessment types (e.g., Lau, 2016; Perera, 2014).

The key findings of this longitudinal study, as elucidated through quantitative data analysis, are summarized as follows:

4.1. Improvement in listening comprehension

The data analysis presented in the previous section indicates that the implementation of summative assessment as the sole method of assessment does not yield significant improvements in learners' performance in listening comprehension. This finding contrasts with the conclusions drawn by Ghahremani (2013). Furthermore, the results indicate that formative assessment alone does not significantly alter learners' performance in listening comprehension compared to the control group. However, the findings suggest that the collaborative use of both assessment methods enhances learners' development in listening comprehension. This integration of both methods within a single group demonstrates a greater effectiveness than the exclusive application of either summative or formative assessments.

The literature highlights the importance of both formative and summative assessments in enhancing language learners' listening and reading skills (Huseynova, 2023; Wijaya, 2022). Teo et al. (2022) emphasize that learners should actively use metacognitive knowledge to manage their listening processes, which is vital for overcoming comprehension challenges. Safa and Motaghi (2021) further support this by noting that metacognitive strategies, such as self-evaluation of listening performance, significantly improve comprehension outcomes. Therefore, language teachers had better utilize study findings to offer diverse assessments and encourage students to develop metacognitive strategies for evaluating their own and peers' listening skills, focusing on various language components and providing holistic feedback (Zandi, 2022).

4.2. Improvement in reading comprehension

The quantitative data revealed that the three experimental groups—those utilizing summative assessment, formative assessment, and a combination of both—outperformed the control group. This underscores the notion that assessment, irrespective of its form, is effective in improving learners' reading comprehension scores.

The present study supports the findings of prior research while also drawing significant conclusions regarding the importance of assessment in language learning. Moreover, this study confirms the effectiveness of assessment methods on both listening and reading comprehension. A comparative analysis of the outcomes for listening and reading comprehension within the summative and formative groups further emphasizes that employing assessment in EFL contexts—regardless of its type—exerts crucial influences on learners' performance. This is a significant finding which is in line with different studies (e.g., Ismail et al., 2022; Mahshanian et al., 2019; Pangestu et al., 2021). For instance, the findings are in line with Ismail et al.'s (2022) study who asserted that contextualization of the different assessment types allows to more personalized language learning experience, based on which, individuals can progress on their own pace and based on their own needs. This is further supported by Mahshanian et al. (2019) who argued in favor of the integration of formative and summative assessment types which can lead to better results on reading skills in educational settings. Meanwhile, Pangestu et al. (2021) referred to the reading comprehension as a skill that can be directed by effective assessment strategies which need effective methods of intervention in order to be fostered. Due to the fact that formative and summative assessment types are essentially linked to each other (Lau, 2016; Pacheco, 2023; Perera, 2014), the integration of both formative and summative assessment types which bring together the best qualities of these two types of assessment are needed to effectively develop the language skills and improved comprehension outcomes.

However, the distinctive nature of these assessment types can affect learners' academic performance and motivation differently, necessitating tailored interventions (Yazidi, 2023). This calls for varied techniques for evaluating performance, allowing for individualized teaching practices (Ortega & Minchala, 2017). For example, Taras (2005) posits starting with summative assessments while incorporating formative elements. Dixson and Worrell (2016) also note that the implementation modes of these assessments can lead to improved educational outcomes.

As quantitatively demonstrated in this study, the integrated application of summative and formative assessment methods suggests that while summative assessment is vital in EFL settings, its benefits are markedly enhanced when used in conjunction with formative assessment. This is further advocated by Pacheco (2023) in language learning contexts, where he acknowledged a synergic formative and summative assessments types.

Overall, the theoretical framework presented by constructivist learning theory which supports providing ongoing feedback on language learners' activities and opportunities to reflect on learning experiences and improve them (Vygotsky, 1978) is extensively used in this study to utilize formative and summative assessments to improve their own listening and reading comprehension process. Meanwhile, as supported by Assessment for Learning, in this study, assessment is not viewed as an end point, but it is regarded as an opportunity to teach, learn, and reflect (Black & Wiliam, 1998; Wiliam, 2011).

5. Conclusion and Implications

This study examined the effects of three assessment methods—summative, formative, and their interaction—on advanced Iranian EFL learners' listening and reading comprehension. The

findings indicate that relying on a single assessment method does not guarantee improvements in listening skills, as both summative-only and formative-only approaches showed insignificant differences. Notably, only the integrated groups made significant progress. In terms of reading comprehension, assessment played a crucial role in facilitating advancement across all groups, highlighting its importance in language learning. While formative assessment is widely supported, the study suggests its benefits are maximized when combined with summative assessment, especially through standardized tests. Given the varied assessment practices among Iranian teachers, further research is needed to enhance assessment literacy in teacher training centers. Such studies could pave the way for more effective integration of assessment types over time. Ultimately, fostering a reflective learning environment with personalized feedback and increased student engagement through integrated approaches is essential.

5.1. Implications

The effectiveness of both summative and formative assessments, especially in integrated formats, suggests that language teachers and curriculum developers should adopt a blended assessment approach. This enhances learners' comprehension and promotes a deeper understanding of language proficiency. The study advocates for varied assessment strategies in EFL programs to address diverse learning needs.

According to the Assessment for Learning approach, both language teachers and learners can view formative and summative assessments as opportunities for improving teaching and learning, rather than just measurement tools. The study emphasizes that assessment should prioritize promoting effective teaching and learning processes. It also supports self-regulated learning, enabling learners to manage their own progress and adjust strategies based on feedback. Thus, these assessment types serve as better indicators of instructional effectiveness, combining teacher monitoring with students' self-regulation.

In spite of the valuable findings of the study, some limitations warrant acknowledgment. The small sample size of advanced Iranian EFL learners restricts the generalizability of results, and the focus on listening and reading may overlook other essential skills like speaking and writing. Future research should include a more diverse participant pool and examine assessment methods across all language domains. Additionally, incorporating qualitative methods could provide deeper insights into learner experiences.

Future studies should also investigate the long-term effects of blended assessment methods on language learning outcomes through longitudinal research. Exploring these methods in various educational contexts and among different proficiency levels would enhance understanding of their effectiveness. Furthermore, examining the role of technology in assessment practices could reveal how digital tools can support formative and summative assessments in EFL settings, leading to a more comprehensive view of assessment's role in language learning and teaching.

Acknowledgements

References

Assessment Reform Group. (2002). *Research-based principles of assessment for learning to guide classroom practice.* London: Nuffield Foundation.

Abdolmajid, M., Mowlaie, B., & Rahimi, A. (2021). Comparative effects of partial dictation and dictogloss on listening comprehension ability of EFL learners. *Global Journal of Foreign Language Teaching,* 11(3), 160-173. https://doi.org/10.18844/gjflt.v11i3.4783

Ahmadi, M., Ismail, H., & Kabilan, M. (2013). The importance of metacognitive reading strategy awareness in reading comprehension. *English Language Teaching,* 6(10). https://doi.org/10.5539/elt.v6n10p235

Al-Qahtani, A. (2015). The effect of explicit instruction of textual discourse markers on Saudi EFL learners' reading comprehension. *English Language Teaching,* 8(4). https://doi.org/10.5539/elt.v8n4p57

Alshahrani, T. (2023). Does the use of web-based extensive listening activities affect foreign language listening anxiety and foreign language anxiety? *Journal of Language Teaching and Research,* 14(6), 1710-1721. https://doi.org/10.17507/jltr.1406.30

Ashraf, H., Fatemi, M., & Naderi, S. (2013). The effect of active learning instruction on the intermediate Iranian EFL learners' listening comprehension ability. *International Journal of Linguistics,* 5(5), 225. https://doi.org/10.5296/ijl.v5i5.4264

Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bayat, A. (2017). The beneficial impact/s of applying formative assessment on Iranian university students' anxiety reduction and listening efficacy. *International Journal of Languages Education,* 5(2), 1-11. https://doi.org/10.18298/ijlet.1740

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2010). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan,* 92(1), 81-90. https://doi.org/10.1177/003172171009200119

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, 22(1), 161-177. https://doi.org/10.1080/0969594X.2014.999643

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215-232. https://doi.org/10.1080/09695941003696016

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469. https://doi.org/10.1080/0969594X.2011.557020

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. https://doi.org/10.1080/0969595980050102

Buck, G. (2001). An overview of listening comprehension. *Assessing listening*, 1-30.

Campione, J. C. (1989). Assisted assessment: a taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities* 22,151–65. https://doi.org/10.1177/002221948902200303

Chen, H. (2020). A contrastive analysis of classroom-based language assessments. *English Language Teaching,* 13(5), 110-114. https://ccsenet.org/journal/index.php/elt/article/download/0/0/42586/44447

Chen, I. J. (2016). Hypertext glosses for foreign language reading comprehension and vocabulary acquisition: Effects of assessment methods. *Computer Assisted Language Learning*, 29(2), 413-426. https://doi.org/10.1080/09588221.2014.983935

Covic, T., & Jones, M. K. (2008). Is the essay resubmission option a formative or a summative assessment and does it matter as long as the grades improve? *Assessment & Evaluation in Higher Education*, 33(1), 75-85. https://doi.org/10.1080/02602930601122928

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, 55(2), 153-159. https://doi.org/10.1080/00405841.2016.1148989

Gan, Z., & Leung, C. (2020). Illustrating formative assessment in task-based language teaching. *ELT Journal,* 74*(1),* 10-19. https://doi.org/10.1093/elt/ccz048

Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms: An Italian case study. *Language Testing*, 17*(2),* 278-288. https://doi.org/10.1177/026553220001700210

Ghahremani, D. (2013). The Effects of Implementing Summative Assessment, Formative Assessment and Dynamic Assessment on Iranian EFL Learners' Listening Ability and Listening Strategy Use. *Journal of Language and Translation 3,* 2*(5),* 59-68. https://journals.iau.ir/article_514741_e0c81dbfe5e918d5aceb6931727e44fe.pdf

Ghiatău, R., Diac, G., & Curelaru, V. (2011). Interaction between summative and formative in higher education assessment: students' perception. *Procedia-Social and Behavioral Sciences*, 11, 220-224. https://doi.org/10.1016/j.sbspro.2011.01.065

Harris, L. (2007). Employing formative assessment in the classroom. *Improving Schools*, 10(3), 249-260. https://doi.org/10.1177/1365480207082558

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21-27. https://doi.org/10.1016/j.stueduc.2012.04.001

Hosseini, H., & Ghabanchi, Z. (2014). The effect of portfolio assessment on EFL learners' reading comprehension and motivation. *English Language Teaching,* 7(5). https://doi.org/10.5539/elt.v7n5p110

Huseynova, F. (2023). Assessment of students' reading comprehension skills in teaching English. https://doi.org/10.5772/intechopen.110600

Ismail, S., Rahul, D., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia,* 12(1). https://doi.org/10.1186/s40468-022-00191-4

Jiang, Y. (2014). Exploring teacher questioning as a formative assessment strategy. *RELC Journal*, 45(3), 287-304. https://doi.org/10.1177/0033688214546962

Kabri, K. (2023). The role of synchronous and asynchronous multimedia in EFL listening comprehension. *NextGen,* 1(2), 44-53. https://doi.org/10.58660/nextgen.v1i2.39

Lau, A. M. S. (2016). 'Formative good, summative bad?'–A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509-525. https://doi.org/10.1080/0309877X.2014.984600

Mahshanian, A., Shoghi, R., & Bahrami, M. (2019). Investigating the differential effects of formative and summative assessment on EFL learners' end-of-term achievement. *Journal of Language Teaching and Research,* 10(5), 1055. https://doi.org/10.17507/jltr.1005.19

Nieminen, J. H., Asikainen, H., & Rämö, J. (2021). Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: a comparison of summative and formative models. *Studies in Higher Education*, 46(7), 1296-1311. https://doi.org/10.1080/03075079.2019.1688282

Ortega, D., & Minchala, O. (2017). Assessing students in an authentic and ongoing manner in the English classroom. *Theory and Practice in Language Studies, 7*(3), 159. https://doi.org/10.17507/tpls.0703.01

Pacheco, A. G. Q. (2022). Assessment of young English-language learners: Formative and summative strategies. *Revista de Lenguas Modernas, 36,* 1-22. https://doi.org/10.15517/RLM.V0I36.48313

Pangestu, A., Hidayah, J., & Apriani, E. (2021). The impacts of explicit and implicit instructions of English connectors on EFL students' reading comprehension. *Linguists Journal of Linguistics and Language Teaching*, 7(1), 23. https://doi.org/10.29300/ling.v7i1.4305

Perera, L., Nguyen, H. O. A., & Watty, K. I. M. (2014). Formative feedback through summative tutorial-based assessments: The relationship to student performance. *Accounting Education*, 23(5), 424-442. https://doi.org/10.1080/09639284.2014.947093

Richards, J. C., & Sandy, C. (2014). *Placement and evaluation package*. Cambridge University Press.

Safa, M., & Motaghi, F. (2021). Cognitive vs. metacognitive scaffolding strategies and EFL learners' listening comprehension development. *Language Teaching Research,* 28(3), 987-1010. https://doi.org/10.1177/13621688211021821

Sönmez, M., & Cetinkaya, F. C. (2022). The Effect of formative assessment on reading comprehension. *International Journal of Assessment Tools in Education*, 9, 88-108. https://doi.org/10.21449/ijate.1104868

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of Educational Studies*, 53(4), 466-478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education*, 9(2), 172-192. https://doi.org/10.1177/1469787408091655

Teo, L., Gani, N., & Ummah, M. (2022). Listening difficulties and strategies in preparing for IELTS listening test using online learning: a case study of intensive English learners. *Teknosastik,* 20(1), 1. https://doi.org/10.33365/ts.v20i1.1367

Torres, J. O. (2019). Positive impact of utilizing more formative assessment over summative assessment in the EFL/ESL Classroom. *Open Journal of Modern Linguistics*, 9(1), 1-11. https://doi.org/10.4236/ojml.2019.91001

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, 31(5), 505-521. https://doi.org/10.1080/02602930600679506

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wijaya, K. (2022). The utilities of online formative assessment in Indonesian EFL learning contexts during covid-19 era. *Journal of Foreign Language Teaching and Learning*, 7(2), 178-198. https://doi.org/10.18196/ftl.v7i2.14249

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Yazidi, R. (2023). Investigating the influence of formative assessment on the learning process in the English language classroom. *Asian Journal of Education and Training,* 9(1), 23-32. https://doi.org/10.20448/edu.v9i1.4540

Yildirim, A., Oscarson, A. D., Hilden, R., & Fröjdendahl, B. (2024). Teaching summative assessment: A curriculum analysis of pre-service language teacher education in Sweden and Finland. *Journal of Teacher Education*, 75(2), 203-218. https://doi.org/10.1177/00224871231214799

Zandi, S. (2022). A qualitative study on EFL teachers' awareness towards the implementation of dynamic assessment. JLS, 5(3), 398-413. https://doi.org/10.25130/jls.5.3.24

Appendices

Appendix A.

Placement Test Specification

Items	Number of items
Listening	8
Vocabulary	10
Grammar	16
Reading	11

Number of Items in Quizzes

Items	Number of items
Listening	5
Vocabulary	8
Grammar	10
Reading	5

Appendix B.

ANCOVA for Listening Comprehension Post-Test in Summative Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.612[a]	2	3.806	5.223	0.017
Intercept	9.963	1	9.963	13.672	0.002
Pre.list.sum	7.412	1	7.412	10.172	0.005
Group	2.432	1	2.432	3.338	0.085
Error	12.388	17	0.729		
Total	2440.000	20			
Corrected Total	20.000	19			

Appendix C.

ANCOVA for Listening Comprehension Post-Test in Formative Group

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	4.930[a]	2	2.465	2.326	0.128
Intercept	22.564	1	22.564	21.287	0.000
Pre.list.form	4.880	1	4.880	4.604	0.047
Group	1.347	1	1.347	1.271	0.275
Error	18.020	17	1.060		
Total	2421.000	20			
Corrected Total	22.950	19			

Appendix D.

ANCOVA for Listening Comprehension Post-Test in Interaction Group

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	10.089a	2	5.044	5.209	.017
Intercept	22.885	1	22.885	23.633	.000
Pre.list.form	6.039	1	6.039	6.236	.023
Group	5.178	1	5.178	5.347	.034
Error	16.461	17	.968		
Total	2603.000	20			
Corrected Total	26.550	19			

Appendix E.

ANCOVA for Reading Comprehension Post-Test in Summative Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	28.923ᵃ	2	14.462	11.805	.001
Intercept	1.841	1	1.841	1.503	.237
Pre.read.sum	20.473	1	20.473	16.712	.001
Group	6.122	1	6.122	4.997	.039
Error	20.827	17	1.225		
Total	3831.000	20			
Corrected Total	49.750	19			

Appendix F.

ANCOVA for Reading Comprehension Post-Test in Formative Group

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	32.764ᵃ	2	16.382	15.659	.000
Intercept	9.712	1	9.712	9.283	.007
Pre.read.form	21.514	1	21.514	20.564	.000
Group	5.436	1	5.436	5.196	.036
Error	17.786	17	1.046		
Total	3887.000	20			
Corrected Total	50.550	19			

Appendix G.

ANCOVA for Reading Comprehension Post-Test in Interaction Group

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	45.408ᵃ	2	22.704	21.693	.000
Intercept	6.445	1	6.445	6.158	.024
Pre.read.inter	21.208	1	21.208	20.264	.000
Group	4.880	1	4.880	4.663	.045
Error	17.792	17	1.047		
Total	4096.000	20			
Corrected Total	63.200	19			