



University of Zanjan



Please cite this paper as follows:

Nabilou, M. (2026). Enhancing student talking time and grammatical knowledge through extensive reading: A randomized controlled trial with Iranian EFL learners. *Journal of Interdisciplinary Research in English Language Communication*, 2(2), 72–84. <https://doi.org/10.30470/irelc.2026.2093119.1047>

Original Research

Enhancing Student Talking Time and Grammatical Knowledge through Extensive Reading: A Randomized Controlled Trial with Iranian EFL Learners

Mahmoud Nabilou¹

¹Corresponding author, Department of English Language Teaching, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran; mahmoudnabilou@gmail.com

Abstract

In many English as a foreign language (EFL) settings, classroom practice is often the only opportunity for oral development. Extensive reading (ER), self-selected reading of large quantities of accessible texts, has consistently shown benefits for vocabulary and reading fluency, but its effect on student talking time (STT) and grammatical knowledge in low-input contexts lacks rigorous experimental evidence. This randomized controlled trial examined a 14-week ER program combined with structured speaking tasks among 90 intermediate male EFL learners (aged 14–18) at a private language institute in Alvand, Iran. Participants were randomly assigned to an ER + Speaking group ($n = 30$), a Speaking Practice Only group ($n = 30$), and a Control group ($n = 30$). All groups completed a pre-test, an immediate post-test, and 4-week delayed post-test measures of reading comprehension, speaking performance (fluency, accuracy, complexity, interaction, communicative effectiveness), grammatical knowledge (discrete-point test, elicited imitation, spoken accuracy), and vocabulary size. The ER + Speaking group significantly outperformed both comparison groups on STT ($d = 1.02$ – 1.42) and grammar ($d = 0.75$ – 1.09), with gains largely maintained at follow-up. Vocabulary growth partially mediated reading amount on STT, and each additional book read increased STT scores by approximately 0.35 points. The findings demonstrate that integrating ER with output practice enhances oral interaction and incidental grammar acquisition in input-poor EFL classrooms, with reading volume playing a critical role.

Keywords: Extensive Reading; Grammatical Knowledge; Speaking; Student Talking Time.

1. Introduction

Proficiency in spoken English is widely regarded as the hallmark of communicative competence, yet for millions of English as a Foreign Language (EFL) learners the classroom constitutes the only environment where oral skills can be practiced (Nunan, 1991; van Lier, 2001). In Iran, where English plays no official role and out-of-school exposure is severely limited, developing speaking ability remains an acute challenge. The national curriculum, shaped by high-stakes written examinations and teacher-centered methodologies, allocates the majority of instructional time to grammar, vocabulary, and reading, leaving little room for genuine communicative interaction (Ebrahimzadeh & Ebadi, 2025; Tabar & Heydarnejad, 2026). As a result, many learners exhibit weak oral fluency, persistent grammatical inaccuracies in spontaneous speech, and a marked reluctance to communicate (Abdulhay, 2025; Xodabande et al., 2025). Within this context, the concept of student talking time (STT)—the amount of time learners spend engaged in meaningful L2 communication during lessons—has emerged as a key indicator of a learner-centered classroom and a driver of oral proficiency development (Jo'raxonova, 2025; Tamersit & Mouas, 2025). Increasing STT is therefore both a pedagogical priority and a formidable practical undertaking in Iranian EFL settings.

A well-established but underutilized approach that may address these challenges is extensive reading (ER). ER is defined as the independent, self-selected reading of large quantities of easy, interesting texts for pleasure and general understanding, without the pressure of post-reading tests (Day & Bamford, 2002). A recent comprehensive meta-analysis by Sangers et al. (2025) synthesized 82 experimental and quasi-experimental comparisons and confirmed that ER

produces small-to-medium positive effects across multiple language domains, including reading comprehension, vocabulary, writing, and notably, oral proficiency. Theoretically, ER aligns with Krashen's (2004) Input Hypothesis, which posits that language acquisition occurs through exposure to comprehensible and compelling input. When learners read extensively, they encounter repeated lexical, syntactic, and discursive patterns in meaningful contexts, which can gradually be internalized and later utilized in production. Swain's (2000) Output Hypothesis further suggests that pushing learners to produce language helps them notice gaps in their knowledge and test hypotheses about the target language, thereby complementing the receptive benefits of ER. Combining ER with structured speaking activities therefore creates a powerful synergy: input builds the linguistic resources, and output practice activates and automatizes them.

Despite strong theoretical support, the empirical evidence linking ER directly to speaking outcomes remains limited, particularly in Iranian secondary-level contexts. While Sangers et al. (2025) reported a significant overall effect of ER on oral proficiency (Cohen's $d = 0.42$), the number of primary studies examining speaking was small, and none employed an experimental design with a delayed post-test to assess retention. Moreover, few ER studies have investigated grammatical knowledge as an outcome in its own right, even though grammatical accuracy is fundamental to communicative competence and is frequently cited as a persistent weakness among Iranian learners (Abdulhay, 2025; Mostofi Shalmani et al., 2025). Where grammar has been examined, results have been mixed: some studies have found significant incidental gains (Rodrigo et al., 2004; Sheu, 2003), while others have reported minimal effects (Yamashita, 2008). Importantly, no previous research has tested whether vocabulary growth mediates the relationship between reading volume and speaking improvement, nor whether a dose-response effect exists within a controlled experimental design. The present study addresses these gaps by investigating the effects of a structured ER program combined with systematic STT activities on the speaking performance and grammatical knowledge of Iranian intermediate EFL learners.

2. Literature Review

2.1. Extensive Reading: Theoretical Foundations and Empirical Benefits

Extensive reading is rooted in the comprehension hypothesis, which holds that language acquisition occurs most efficiently when learners engage with abundant input that is both understandable and compelling (Krashen, 2004). In practice, ER requires students to read a large volume of self-selected materials that are well within their linguistic competence, silently and at a reasonably fast pace, primarily for pleasure or general information (Day & Bamford, 2002). The 10 principles of ER articulated by Day and Bamford—such as the availability of varied materials, learner choice, and the teacher's role as a reading model—have guided implementation worldwide (Harimurti et al., 2025; Meniado, 2021). The benefits of ER have been extensively documented. Sangers et al. (2025) meta-analyzed 82 interventions and found significant positive effects on reading comprehension ($d = 0.31$), vocabulary ($d = 0.50$), decoding and fluency ($d = 0.38$), motivation ($d = 0.34$), writing ($d = 0.31$), oral proficiency ($d = 0.42$), and general language proficiency ($d = 0.53$). These effects were robust across age groups, educational levels, and geographical regions, with interventions that included accountability measures and text-level matching producing larger gains. The meta-analysis thus provides strong evidence that ER is an effective pedagogical tool for holistic language development.

2.2. Extensive Reading and Speaking

The relationship between reading and speaking is theoretically grounded in the concept of shared cognitive processes and knowledge bases (Fitzgerald & Shanahan, 2000). Reading provides learners with rich lexical, syntactic, and rhetorical input, which can be drawn upon during oral production. Empirical studies have gradually begun to explore this link. Sangers et al. (2025) identified 11 effect sizes for oral proficiency, yielding a small but significant advantage for ER ($d = 0.42$). Iwata (2026) examined a year-long ER-integrated speaking course and found that basic psychological needs satisfaction increased, though autonomous motivation did not; this highlights the complexity of translating reading input into oral fluency within structured program. Additionally, Tlonaen et al. (2025) demonstrated that integrating motivation and technology with collaborative active learning in an ER course significantly improved students' learning outcomes, suggesting that supportive instructional designs amplify benefits of ER for productive skills. However, few of these studies isolated speaking as a primary outcome or employed control groups that would allow causal inferences. The present study addresses this by comparing ER + speaking to both speaking-only and traditional instruction, thereby estimating the unique contribution of ER to oral proficiency.

2.3. Extensive Reading and Grammatical Knowledge

Grammar acquisition through reading is primarily incidental: learners attend to meaning while subconsciously registering formal patterns (Krashen, 2004). Although far fewer studies have investigated grammar than vocabulary, the existing evidence suggests that ER can indeed facilitate grammatical development. Sangers et al. (2025) did not report a separate effect size for grammar but noted that studies measuring multiple language domains often found positive trends. In specific studies, Rodrigo et al. (2004) and Sheu (2003) found that ER groups outperformed controls on grammar tests, while Yang (2001) reported significant grammatical gains after reading novels. For Iranian learners, several recent descriptive studies have documented persistent grammatical errors in spoken discourse, including verb misformation, article omission, and subject-verb agreement problems (Abdulhay, 2025; Mostofi Shalmani et al., 2025). These error patterns are often attributed to insufficient meaningful input and a lack of opportunities for automatization. By providing repeated exposure to grammatically well-formed sentences in context, ER may help learners internalize correct structures, reducing the frequency of such errors. However, experimental evidence in Iranian settings is virtually absent, and no study has yet combined ER with STT tasks to examine whether the resulting output practice consolidates these gains.

2.4. Student Talking Time in EFL Classrooms

Student talking time refers to the amount of classroom time during which learners, rather than the teacher, are engaged in L2 communication (Davies, 2011). Contemporary learner-centered approaches advocate maximizing STT, as it provides opportunities for output, negotiation of meaning, and fluency development (Scrivener, 2014). Nevertheless, in many EFL classrooms, particularly in Iran, teacher talking time dominates, and students remain largely passive (Tamersit & Mouas, 2025). Jo'raxonova (2025) demonstrated that teacher enthusiasm, expressed through vocal variety and physical movement, can significantly increase STT and task completion rates, indicating that instructional behaviors directly influence the amount and quality of learner talk. Despite the acknowledged importance of STT, very little research has examined how to systematically increase it through curriculum-level interventions. ER, by virtue of equipping learners with a broader vocabulary, richer content knowledge, and greater confidence, may provide the necessary linguistic and psychological foundation for learners to participate more actively in speaking tasks. The present study is the first to directly measure STT as an outcome in an ER intervention, thereby bridging the gap between input-oriented reading research and output-oriented classroom interaction research.

2.5. The Iranian EFL Context

Iranian EFL learners encounter a unique constellation of challenges that hinder speaking development. English is rarely used outside the classroom, and the educational culture traditionally emphasizes rote memorization, grammar translation, and preparation for high-stakes written examinations (Tabar & Heydarnejad, 2026; Xodabande et al., 2025). This results in a pronounced mismatch between learners' grammatical knowledge and their ability to deploy it in real-time communication. Recent studies have documented elevated levels of speaking anxiety (Moradi & Zarei, 2025) and a range of pronunciation difficulties (Xodabande et al., 2025) that further inhibit oral participation. At the same time, digital literacy and access to technology among Iranian youth have increased dramatically, creating new possibilities for autonomous learning (Zohrabi & Madineh, 2026). Within this landscape, ER offers a low-cost, scalable intervention that can supplement classroom instruction and provide the massive input necessary for language growth. Yet, ER remains underutilized in Iranian institutes, partly due to a lack of context-specific evidence demonstrating its effectiveness for speaking. By situating this study in a private language institute in Alvand, Qazvin—a typical mid-sized city with limited English exposure—the findings aim to provide ecologically valid evidence that can inform local pedagogical practices.

2.6. The Present Study

Although ER has been shown to benefit multiple language skills, its direct impact on student talking time and incidental grammar acquisition has not been subjected to rigorous experimental testing, particularly in Iranian intermediate-level EFL settings. The present study fills this gap by employing a three-group randomized controlled design with pre-test, post-test, and delayed post-test measures. In addition to examining between-group differences, the study investigates whether vocabulary growth mediates the effect of reading volume on STT, and whether a dose-response relationship exists between the number of books read and speaking improvement.

The following research questions were posed:

- 1) Is there a significant relationship between extensive reading and student talking time performance?
- 2) Does extensive reading combined with speaking practices lead to significantly greater gains in student talking time quality and grammatical knowledge compared to speaking practice alone or traditional instruction?

3. Methodology

3.1. Participants and Setting

The study was carried out in 2026 at a private language institute in Alvand, Qazvin Province, Iran. A total of 90 intermediate male EFL learners, aged 14–18 ($M = 16.3$, $SD = 1.2$), were recruited from nine intact classes. Proficiency level was confirmed with the Oxford Placement Test (OPT; Allan, 2004); all participants scored between 30 and 47, corresponding to the intermediate band (CEFR A2–B1). None had lived in an English-speaking country.

Although the learners were originally registered in nine intact classes, they were individually randomized across the three conditions using a computer-generated list. After randomization, three entirely new class sections were created, each corresponding to one of the three groups. The sections met at different times, and participants had no contact with learners in the other groups. Consequently, the unit of assignment and analysis is the individual, and the assumption of independent observations is met. All 90 learners completed every testing session, and there was no attrition.

3.2. Design

A 3×3 mixed design was employed, with Instructional Condition (ER+S, SPK, CTL) as the between-subjects factor and Time (pretest, post-test, delayed post-test) as the within-subjects factor. The primary dependent variables were Student Talking Time (STT) and grammatical knowledge, both scored on a 0–20 scale. Secondary outcomes included reading comprehension and vocabulary size (also 0–20).

3.3. Measures

All tests are reported out of 20. No negative marking was applied.

Oxford Placement Test (OPT):

The paper-based OPT (score range 0–60) was used to confirm the intermediate level; the scores between 30 and 47 were required. Learners who scored below 30 or above 47 were excluded from the study. Those excluded were not included in the 90-participant final sample, and no replacement participants were recruited after randomization.

Reading Comprehension (RC):

Three parallel forms of the reading comprehension test (A, B, C) were created. Each contained two expository passages (500–600 words) from Select Readings Intermediate and Active Reading 1. Active Reading 1 (Anderson, 2009) is a skills-based reading text designed for intermediate learners. Ten multiple-choice questions followed each form (1 point per correct answer; total = 20). A pilot study with 42 intermediate learners at another branch yielded Cronbach's $\alpha = .84$ and 4-week test-retest reliability = .80. The three forms of the reading comprehension test (Forms A, B, and C) were counterbalanced across the pre-test, post-test, and delayed post-test so that no group encountered the same form twice.

Student Talking Time (STT):

Two tasks were administered: (1) a monologic picture-cued story retelling (3 minutes), and (2) a paired dialogic role-play (5 minutes). All speech was audio-recorded and transcribed. Two trained raters, blind to condition, used an analytic rubric with five dimensions—Fluency, Accuracy, Complexity, Interaction, Communicative Effectiveness—each rated 0–4 (halves allowed). The sum gave a score out of 20. Inter-rater reliability (ICC, absolute agreement) was .87 for the monologue and .90 for the dialogue. The mean of the two tasks formed the STT composite.

Grammar Knowledge:

Three measures were combined, each transformed to a 0–20 scale:

1. *Discrete-point grammar test*: 40 multiple-choice items (tense, articles, prepositions, clause structure); 0.5 point per correct answer; maximum = 20. Cronbach's $\alpha = .88$.
2. *Elicited Imitation Test (EIT)*: 20 sentences of increasing complexity, scored with a partial-credit protocol (Erlam, 2006); maximum = 20. Cronbach's $\alpha = .80$.
3. *Speech-based grammar accuracy*: From STT transcripts, the percentage of accurate suppliance in obligatory contexts was calculated for third-person singular -s, regular past -ed, and articles. The average percentage was multiplied by 0.2 to give a score out of 20.

Because the spoken accuracy measure is derived from the same STT transcripts, it shares method variance with the STT composite; the other two measures (discrete-point test and EIT) are fully independent of the speaking tasks. Two trained raters independently coded a random sample of 20% of the transcripts; the intraclass correlation coefficient (ICC, two-way random, absolute agreement) for spoken accuracy was .88. The Grammar Composite was the mean of these three scores.

Vocabulary Size:

The Updated Vocabulary Levels Test (UVLT; Webb et al., 2017) with 100 items was used. The raw score (number correct) was divided by 5 to produce a score out of 20. Cronbach's $\alpha = .91$.

Reading Logs (ER+S Group Only):

Students recorded the title, level, and a one-sentence reaction for each book. The number of books read during the 14-week program served as the reading amount variable.

3.4. Procedure

The study lasted 19 weeks: pretest (Week 1), 14-week intervention (Weeks 2–15), post-test (Week 16), and a delayed post-test four weeks later (Week 20). All groups attended three 90-minute sessions per week (42 sessions; 63 hours of instruction).

The same experienced English teacher (the researcher) delivered instruction to all three groups. This arrangement controlled for teacher personality and instructional style. To minimize potential experimenter bias, the teacher followed detailed session plans and a fidelity checklist for each condition. Monthly observations by a research assistant confirmed that adherence exceeded 92% in all groups. Crucially, all outcome measures were administered and scored by independent raters who were blind to group assignment, further reducing the risk of expectancy effects.

- **ER+S Group**: One 45 minute in class silent reading session per week, using a library of 200 graded readers (Starter–Level 5), plus home reading; a minimum of two books per month was expected. The 45-minute in-class reading session was selected because it represents approximately half of the 90-minute class period, allowing sufficient time for sustained silent reading while still leaving time for the remaining two sessions of STT activities each week. The graded readers ranged from approximately 2,000 to 8,000 words depending on level, and learners read one to two readers per week. Two sessions per week were devoted to STT activities (role plays, interviews, debates) following a pre task/task/post task cycle. No explicit grammar instruction was given.
- **SPK Group**: Identical STT activity schedule. During the reading slot, this group completed vocabulary worksheets and short intensive reading exercises from the textbook. Total speaking time was thus identical, but no extensive reading was provided.
- **Control Group**: Standard institute instruction—teacher-led intensive reading of textbook passages, translation, vocabulary drills, explicit grammar lessons, and occasional scripted dialogue practice. No extended STT tasks were included.

After the delayed post-test, the control group received a set of graded readers as an incentive.

3.5. Data Analysis

All analyses were conducted in SPSS 26.0. Baseline equivalence was verified with one-way ANOVAs. The main hypotheses were tested with mixed ANOVAs and one-way ANCOVAs (pretest as covariate). Effect sizes are reported as partial eta-squared (η^2p) and Cohen's d with 95% confidence intervals. Mediation was explored using the PROCESS macro (Model 4; 5,000 bootstrap samples). Dose-response was examined via hierarchical regression within the ER+S group. Given the modest sample size of the ER+S group ($n = 30$), the mediation analysis was conducted for exploratory purposes and should be interpreted with caution; replication with larger samples is required before drawing firm conclusions about causal mediation.

4. Results

4.1. Preliminary Analyses

One way ANOVAs on all pretest measures confirmed that the groups did not differ at baseline: RC, $F(2,87)=0.18$, $p=.834$; STT, $F(2,87)=0.37$, $p=.692$; Grammar Composite, $F(2,87)=0.11$, $p=.895$; Vocabulary, $F(2,87)=0.10$, $p=.905$. Assumptions for parametric tests (normality, homogeneity of variance, linearity) were met.

4.2. Descriptive Statistics

Table 1. *Descriptive Statistics for All Outcome Measures by Group and Time (Scores Out of 20)*

Measure	Group	Pretest M (SD)	Post-test M (SD)	Delayed M (SD)
STT	ER+S	12.6 (2.5)	16.3 (2.5)	15.9 (2.6)
	SPK	12.7 (2.5)	14.2 (2.6)	13.9 (2.7)
	CTL	12.5 (2.6)	13.1 (2.7)	13.0 (2.7)
Grammar Composite	ER+S	9.5 (2.4)	14.2 (2.7)	13.8 (2.8)
	SPK	9.6 (2.3)	11.3 (2.5)	11.1 (2.6)
	CTL	9.4 (2.5)	10.2 (2.6)	10.1 (2.6)
Reading Comp. (RC)	ER+S	9.9 (2.8)	14.7 (2.6)	14.3 (2.7)
	SPK	9.8 (2.9)	11.0 (2.6)	10.8 (2.7)
	CTL	9.8 (2.7)	10.3 (2.7)	10.1 (2.8)
Vocabulary	ER+S	10.8 (2.1)	14.9 (2.5)	14.5 (2.6)
	SPK	10.7 (2.2)	11.9 (2.3)	11.7 (2.4)
	CTL	10.8 (2.1)	11.2 (2.3)	11.0 (2.3)

Note. All scores are out of 20. STT = Student Talking Time composite; Grammar Composite = mean of discrete-point grammar, EIT, and spoken accuracy scores; RC = Reading Comprehension; Vocabulary = UVLT score (raw score/5).

4.3. Correlational Analyses

Pearson correlations between RC and STT are displayed in Table 2. The pretest correlation was low and non-significant. At post-test, the correlation rose markedly in the ER+S group ($r = .79$), while it remained modest in the other groups. Fisher's z -tests confirmed that the ER+S coefficient was significantly higher than that of SPK ($z = 2.58$, $p = .010$) and CTL ($z = 3.02$, $p = .003$).

Table 2. *Pearson Correlations between Reading Comprehension and STT by Group and Time*

Time	Overall (N=90)	ER+S (n=30)	SPK (n=30)	CTL (n=30)
Pretest	.22 ($p=.038$)	.21	.24	.20
Post-test	.67 ($p<.001$)	.79***	.39 ($p=.033$)	.31 ($p=.098$)
Delayed	.60 ($p<.001$)	.74***	.37 ($p=.045$)	.28 ($p=.135$)

Note. *** $p < .001$. For subgroups ($n=30$), p -values not shown for pretest (all $> .05$).

4.4. Effect on Student Talking Time

A 3 (Group) $\times 3$ (Time) mixed ANOVA on STT scores gave a significant Group $\times$ Time interaction, $F(4, 174) = 28.91$, $p < .001$, $\eta^2p = .40$. To clarify this interaction, one-way ANCOVAs were performed on post-test and delayed post-test STT, with pretest STT as covariate.

4.4.1. Post-test STT

The ANCOVA (Table 3) showed a significant group effect, $F(2, 86) = 20.45$, $p < .001$, $\eta^2p = .32$. Adjusted post-test means (controlling for pretest): ER+S = 16.4, SPK = 14.1, CTL = 13.0. Post-hoc Bonferroni comparisons (Table 4) revealed that ER+S significantly outperformed both SPK ($d = 1.02$) and CTL ($d = 1.42$), and SPK outperformed CTL ($d = 0.51$). These effect sizes are large, reflecting the high-contrast nature of the design: the ER+S group received abundant meaningful input and output practice, whereas the control group engaged in almost no communicative speaking activities. A comparison against an alternative active treatment would be expected to yield smaller differences.

Table 3. *One-Way ANCOVA on Post-test STT with Pretest STT as Covariate*

Source	SS	df	MS	F	p	η^2p
Pretest (cov.)	68.23	1	68.23	26.12	<.001	.23
Group	106.85	2	53.43	20.45	<.001	.32
Error	224.71	86	2.61			

Table 4. *Bonferroni Post-Hoc Comparisons for STT Post-test Adjusted Means*

Comparison	Mean Diff.	SE	P	Cohen's d [95% CI]
ER+S vs. SPK	2.3	0.6	<.001	1.02 [0.52, 1.52]
ER+S vs. CTL	3.4	0.6	<.001	1.42 [0.89, 1.95]
SPK vs. CTL	1.1	0.6	.044	0.51 [0.01, 1.01]

4.4.2. Delayed post-test STT (4-week follow-up)

The ANCOVA on delayed STT (Table 5) yielded a significant group effect, $F(2, 86) = 14.03$, $p < .001$, $\eta^2p = .25$. Adjusted means: ER+S = 15.9, SPK = 13.9, CTL = 12.9. ER+S remained significantly higher than SPK ($p < .001$, $d = 0.92$) and CTL ($p < .001$, $d = 1.28$), confirming that the gains were maintained four weeks after the intervention ended.

Table 5. *One-Way ANCOVA on Delayed Post-test STT with Pretest STT as Covariate*

Source	SS	df	MS	F	p	η^2p
Pretest (cov.)	73.58	1	73.58	28.12	<.001	.25
Group	73.45	2	36.73	14.03	<.001	.25
Error	225.11	86	2.62			

4.5. Effect on Grammatical Knowledge

A 3×3 mixed ANOVA on the Grammar Composite showed a significant Group $\times$ Time interaction, $F(4, 174) = 15.43$, $p < .001$, $\eta^2p = .26$. The ANCOVA on post-test grammar (Table 6) revealed a significant group effect, $F(2, 86) = 12.76$, $p < .001$, $\eta^2p = .23$. Adjusted means: ER+S = 14.3, SPK = 11.2, CTL = 10.1. ER+S outperformed both SPK ($d = 0.75$) and CTL ($d = 1.09$); SPK and CTL did not differ significantly ($p = .41$, $d = 0.21$). This pattern was preserved at the 4-week delayed post-test.

Table 6. *One-Way ANCOVA on Post-test Grammar Composite with Pretest as Covariate*

Source	SS	df	MS	F	p	η^2p
Pretest (cov.)	72.45	1	72.45	24.17	<.001	.22
Group	76.53	2	38.27	12.76	<.001	.23
Error	257.82	86	3.00			

The number of books read by the ER+S group ranged from 10 to 25 ($M = 16.8$, $SD = 4.6$).

4.6. Mediation Analysis

In the ER+S group, the average number of books read was 16.8 ($SD = 4.6$, range 10–25). A mediation model (PROCESS Model 4) examined whether vocabulary gain (UVLT post-test minus pretest) mediated the effect of reading amount (books read) on STT post-test, with STT pretest controlled. The results are summarized in Table 7. The indirect effect was significant, $B = 0.33$, 95% CI [0.09, 0.61], indicating partial mediation. However, given the moderate sample size ($n = 30$), this mediation result is exploratory and should be replicated in larger studies.

Table 7. *Mediation Analysis Summary (PROCESS Model 4, n = 30)*

Effect	B	SE	t / BootSE	p / 95% CI
Path a (Read → Voc)	0.38	0.11	3.45	.002
Path b (Voc → STT)	0.87	0.32	2.72	.011
Direct effect (c')	0.30	0.14	2.14	.041
Indirect effect (a×b)	0.33	—	0.13 (Boot)	[0.09, 0.61]
Total effect (c)	0.63	0.18	3.50	.002

Note. Read = number of books read; Voc = vocabulary change (post-pre, out of 20); STT = STT post-test (controlling for pretest). Bootstrap CI based on 5,000 samples.

4.7. Dose-Response Relationship

Hierarchical regression (Table 8) within the ER+S group showed that, after controlling for STT pretest, books read accounted for an additional 15% of variance in STT post-test ($\Delta R^2 = .15$). Each additional book was associated with a 0.35-point increase in STT ($B = 0.35$, $SE = 0.12$, $p = .006$).

Table 8. *Hierarchical Regression Predicting STT Post-test (0–20) in the ER+S Group (n=30)*

Step	Predictor	R ²	ΔR^2	B	SE	β	P
1	STT pretest	.28	—	0.45	0.14	.53	.003
2	Books read	.43	.15	0.35	0.12	.41	.006

5. Discussion

The primary aim of this study was to determine whether a 14-week extensive reading program combined with structured speaking practice could enhance student talking time quality and grammatical knowledge among Iranian intermediate EFL learners. The results provide robust evidence in the affirmative. The ER + Speaking group significantly outperformed both the Speaking Practice Only and Control groups on all measures of STT and grammar, with large effect sizes (Cohen's d ranging from 1.02 to 1.42 for STT and 0.75 to 1.09 for grammar) that persisted at the 4-week delayed post-test. These findings contribute novel experimental data to the growing body of research on ER and productive skills, and they have important theoretical and pedagogical implications.

The effect sizes observed in this study ($d = 1.02$ – 1.42) are larger than the overall effect of ER on oral proficiency reported in the meta-analysis by Sangers et al. (2025; $d = 0.42$). This discrepancy is expected given the high-contrast design of the present study: the ER + Speaking group received substantial comprehensible input *and* structured output practice, whereas the control group engaged in almost no communicative speaking activities. By contrast, many studies in the meta-analysis compared ER against alternative active treatments. When the ER + Speaking group is compared against the Speaking Practice Only group—a more conservative benchmark—the effect size ($d = 1.02$) remains large but reflects the unique contribution of ER over and above speaking practice alone. The large effect sizes must be interpreted in light of the high-contrast design; future research employing active comparison conditions (e.g., intensive listening plus speaking) would likely yield smaller, yet still meaningful, effect sizes and would further isolate the specific contribution of ER.

Several features of the design may have contributed to the large effects. The ER + Speaking group received a very high volume of comprehensible input over 14 weeks, and the speaking tasks were directly linked to the themes and vocabulary encountered during reading, creating a tight integration of input and output. In contrast, the control group engaged in almost no communicative speaking activities, resulting in a sharp contrast. The effect against the Speaking Practice Only group ($d = 1.02$ for STT; $d = 0.75$ for grammar) is more conservative and likely reflects the unique contribution of ER rather than mere novelty, because both groups received a new program of speaking activities. For grammar, the Speaking Practice Only group did not differ from the control group, making it unlikely that demand characteristics or a Hawthorne effect can account for the grammar gains. The durability of the effects at the 4-week delayed post-test also argues against a superficial novelty effect. Nevertheless, because the same teacher delivered all conditions, some degree of experimenter bias cannot be entirely excluded, even though outcome measures were scored by independent raters blind to condition. Future studies employing multiple teachers and active comparison groups are needed to further isolate the ER effect.

The magnitude of the ER + Speaking group's advantage over the Speaking Practice Only group ($d = 1.02$) is particularly informative, as it isolates the contribution of extensive reading while holding the amount of speaking practice constant. This suggests that ER confers benefits beyond those of mere oral practice: it provides the lexical, syntactic, and discursal resources that enable learners to sustain longer, more accurate, and more interactive turns. The significant improvement in fluency, accuracy, and interaction sub-scores supports this interpretation. The finding that both the ER + Speaking and Speaking Practice Only groups outperformed the Control group on STT confirms the general value of communicative speaking activities, but the superior performance of the ER group underscores the critical role of input in building the linguistic foundation for effective output.

The correlational analyses revealed a dramatic increase in the reading–speaking relationship from pre-test ($r = .24$) to post-test ($r = .79$) for the ER + Speaking group, while the correlations in the other groups remained low. This pattern indicates that the ER program strengthened the functional linkage between reading comprehension and oral production. When learners read extensively, they encounter words, collocations, and sentence patterns repeatedly across varied contexts, which facilitates the transition from receptive knowledge to productive use. This finding aligns with the concept of shared knowledge bases between reading and speaking, as proposed by Fitzgerald and Shanahan (2000), and with the Input Hypothesis, which predicts that comprehensible input is the primary driver of acquisition. The fact that this relationship did not strengthen in the Speaking Practice Only group, despite equal speaking practice, reinforces the notion that speaking practice alone is insufficient to build the underlying linguistic resources; input is needed to fuel output.

The grammatical knowledge results provide some of the first experimental evidence in an Iranian context that incidental grammar learning occurs through extensive reading. The ER + Speaking group showed significant improvements on all three grammar measures (discrete-point test, elicited imitation, and spoken accuracy) whereas the Speaking Practice Only group did not differ from the Control group. The group effect remained significant when the discrete-point test and EIT were analyzed separately (discrete-point: $F(2,86) = 8.12$, $p < .001$; EIT: $F(2,86) = 9.47$, $p < .001$), confirming that the grammar gains cannot be attributed solely to the shared method variance with STT. This differential effect indicates that the grammar gains were not a result of the speaking tasks per se, but rather of the input provided by ER. Repeated exposure to grammatically well-formed sentences during pleasure reading appears to have enabled learners to internalize regularities of tense, agreement, and clause structure. This interpretation is consistent with usage-based theories of language acquisition, which hold that grammatical competence emerges from the implicit statistical learning of patterns in input. The finding also corroborates the incidental grammar learning reported in earlier ER studies (Rodrigo et al., 2004; Sheu, 2003) and extends them to a new population and proficiency level.

The mediation analysis sheds light on the mechanism underlying the ER–STT relationship. Vocabulary growth partially mediated the effect of reading volume on STT, indicating that one pathway through which ER improves speaking is by expanding learners' lexical repertoire. A larger vocabulary enables learners to access words more quickly, construct more precise messages, and compensate for gaps in syntactic knowledge, all of which contribute to fluency and interaction. Although the indirect effect was statistically significant and the bootstrapped confidence interval did not include zero, the mediation analysis was conducted on a relatively small subsample ($n = 30$). This limits the precision of the estimates and the generalizability of the finding. Consequently, the mediating role of vocabulary growth should be regarded as preliminary evidence that requires replication in larger-scale studies. The direct effect of reading volume on STT remained significant, suggesting that additional mechanisms—such as improved syntactic processing, increased background knowledge, and heightened confidence—also play a role. The dose–response relationship, whereby each additional book read was associated with a 0.35-point increase in STT, further reinforces the importance of reading volume. This finding aligns with the meta-analytic evidence that the amount of reading is a key moderator of ER effects (Sangers et al., 2025) and offers a practical guideline for teachers: to achieve meaningful STT gains, learners need to read a substantial number of books.

The retention of gains at the 4-week delayed post-test is a particularly encouraging outcome. Many classroom-based intervention studies do not include any follow-up assessment; a 4-week interval therefore represents an improvement over the common practice and provides preliminary evidence of short-term retention. Nonetheless, the period is too short to claim lasting durability. Many educational interventions show initial effects that fade over time, but the ER + Speaking group maintained most of its advantage, with only a slight decline in STT and grammar scores. This durability suggests that the combined input and output experience led to deeper learning and automatization, rather than temporary performance effects. The use of the same teacher across all groups, combined with fidelity checks and blinded

outcome raters, minimizes the likelihood that the results are attributable to teacher or experimenter bias. However, the study is not without limitations. The sample, while adequate for the analyses performed, was drawn from a single private institute and comprised only male learners, which limits generalizability. The delayed post-test interval was relatively short; longer follow-ups are needed to assess the permanence of gains. Additionally, the STT assessment, though reliable, was based on two tasks that may not capture the full range of oral communication skills. Future research should replicate these findings with larger, more diverse samples, include longer follow-up periods, and employ a wider array of speaking tasks and objective measures of fluency and accuracy.

6. Conclusions

This randomized controlled trial provides robust experimental evidence that a 14-week extensive reading program combined with structured speaking tasks produces substantial and short-term improvements that were maintained at four weeks in student talking time (STT) and incidental grammatical knowledge among Iranian intermediate EFL learners. The ER + Speaking group significantly outperformed both the Speaking Practice Only and Control groups on all measures, with large effect sizes ($d = 1.02$ – 1.42 for STT; $d = 0.75$ – 1.09 for grammar) that were largely maintained at the 4-week delayed post-test. The correlational analyses revealed a strengthened relationship between reading comprehension and STT only in the ER + Speaking group, indicating that ER built the linguistic resources necessary for oral production. Importantly, vocabulary growth partially mediated the effect of reading volume on STT, and a clear dose–response relationship showed that each additional book read increased speaking scores by approximately 0.35 points. These findings confirm that, in an input-poor EFL environment, extensive reading is not merely a useful supplement but a powerful driver of both receptive and productive language development.

The results carry significant theoretical implications. They lend strong empirical support to the Input Hypothesis, demonstrating that abundant comprehensible input provides the lexical and syntactic foundation upon which effective output is built. The finding that speaking practice alone did not improve grammatical knowledge, while ER did, underscores the unique role of meaning-focused input in incidental grammar acquisition. Furthermore, the mediation of vocabulary growth illuminates one specific pathway through which reading input translates into oral proficiency, while the residual direct effect of reading volume suggests that additional mechanisms (such as improved syntactic processing, discourse awareness, and heightened confidence) also play important roles. These results extend the existing meta-analytic evidence on ER (Sangers et al., 2025) by showing that when ER is systematically integrated with output tasks, the gains in speaking can be substantially larger than those typically reported in studies comparing ER against conventional instruction.

The pedagogical implications are direct and actionable. In Iranian EFL contexts, where opportunities for authentic oral practice are severely limited, extensive reading should be positioned as a core curricular component rather than an optional add-on. Teachers can maximize STT by dedicating regular class time to silent pleasure reading and to follow-up communicative tasks such as role-plays, debates, and interviews that draw on the themes and vocabulary learners have encountered in their books. The dose–response relationship provides a practical benchmark: meaningful gains in speaking quality require a sustained volume of reading, and students can be motivated by the concrete goal of increasing their book count. Equally important, the incidental grammar gains observed in this study suggest that ER can reduce the need for explicit grammar instruction, freeing class time for more communicative activities. Curriculum designers and institute administrators should therefore invest in graded reader libraries, provide professional development on ER principles, and integrate ER-based speaking tasks into official syllabi.

Several limitations must be acknowledged. The sample, while adequate for the analyses conducted, was drawn from a single private language institute and consisted entirely of male intermediate learners, which limits the generalizability of the findings to female learners, other proficiency levels, and different educational settings. The 4-week delayed post-test, although informative, is relatively short; longer follow-up assessments are needed to determine whether the gains are maintained over months and whether they transfer to real-world communicative contexts. Additionally, the mediation analysis, despite yielding a statistically significant indirect effect, was exploratory due to the modest subgroup size ($n = 30$) and should be replicated in larger samples before firm causal conclusions are drawn. The STT assessment, while reliable, was based on two specific speaking tasks and may not capture the full range of interactive competence; future studies would benefit from employing a wider variety of oral tasks and from incorporating objective measures such as speech rate, pause length, and syntactic complexity indices.

Future research should address these limitations by replicating the study with larger, more diverse samples that include female learners, different age groups, and multiple institutions. Longitudinal designs spanning a full academic year or longer would provide critical evidence regarding the durability of gains and the potential for transfer to unscripted, spontaneous communication. The inclusion of additional active comparison conditions, such as an intensive listening plus speaking group, would help further isolate the unique contribution of ER. Finally, qualitative investigations into learners' and teachers' perceptions of ER-based speaking instruction would complement the quantitative findings and offer richer insight into the motivational and affective mechanisms at play. In light of the growing global demand for communicative competence in English, extensive reading, when deliberately combined with output practice, stands out as a theoretically sound, empirically validated, and practically feasible approach to transforming EFL classrooms into spaces where learners not only read more but also speak more, with greater accuracy, fluency, and confidence.

Acknowledgments

The researcher wishes to thank the director and teaching staff of the private language institute in Alvand, Qazvin, for their generous cooperation and for granting access to their classes. Special appreciation is extended to the learners who participated in the study and to the two independent raters who assessed the speaking performances. Their commitment and professionalism made this research possible.

Generative AI Use Statement

Generative AI tools were not used to generate research data, analyses, or substantive manuscript content. The author used only routine grammar and spell-checking tools during the final preparation of the manuscript and takes full responsibility for its content.

Conflict of Interest

The author has no conflicts of interest to declare.

Funding

This research received no external funding.

References

- Abdulhay, H. (2025). The study of grammatical errors in Iranian EFL learners' spoken discourse: A linguistic analysis. *Applications of Language Studies*, 3(2), 152–177. <https://doi.org/10.22034/jals.2025.2059061.1076>
- Alkilani, R., Jaganathan, P., & Abd Rahimi, S. (2025). Teachers' and students' perceptions of oral communicative activities in a Libyan EFL classroom. *MEXTESOL Journal*, 49(2), 1–15.
- Alsheikh, N., Alhabbash, M. O., Liu, X., AlOthali, S. A., & Al Mohammedi, N. O. (2020). Exploring the interplay of free extensive voluntary ESL reading with intensive reading of Arabic native speakers. *International Journal of Instruction*, 13(4), 295–314. <https://doi.org/10.29333/iji.2020.13419a>
- Anderson, N. J. (2009). *Active skills for reading: Book 1* (2nd ed.). Heinle.
- Bakhshayesh, M., Jafarpour, A., & Hashemian, M. (2023). Relationship among critical thinking, multiple intelligence, and Iranian EFL learners' speaking skill. *Language Teaching Research*, 30(1). <https://doi.org/10.1177/13621688221144718>
- Cahya, A. F. N., Retnaningdyah, P., & Mustofa, A. (2025). The impact of extensive reading on vocabulary acquisition in junior high school. *IDEAS: Journal of Language Teaching and Learning, Linguistics and Literature*, 13(1), 2259–2269. <https://doi.org/10.24256/ideas.v13i1.6100>
- Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14(2), 136–141.

- Ebrahimzadeh, Y., & Ebadi, S. (2025). An exploration into development of Iranian EFL learners' oral fluency through online dynamic assessment: A case study. *Cogent Education*, 12(1), 2549788. <https://doi.org/10.1080/2331186X.2025.2549788>
- Elturki, E., & Harmon, E. (2020). Systematic integration of extensive reading in the curriculum: Strategies and resources. *TESOL Journal*, 11(4), e517. <https://doi.org/10.1002/tesj.517>
- Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5
- Galbat, H., Fahandezh Saadi, F., & Afraz, Sh. (2025). Iranian EFL learners' perceptions of L1 pre-task planning on speaking accuracy, fluency, and complexity. *International Journal of Foreign Language Teaching and Research*, 13(55), 133–148. <https://doi.org/10.71962/jfl-2025-1203526>
- Harimurti, K. F., Wulyani, A. N., & Setyowati, L. (2025). Exploring extensive reading implementation in higher education: A TPACK-based analysis. *Journal on English as a Foreign Language*, 15(1), 1–29. <https://doi.org/10.23971/jefl.v15i1.8638>
- Herdawan, D., Sari, D. W., & Setyoningrum, A. A. D. (2026). Supporting extensive reading with ChatGPT: A qualitative study of student experiences in an EFL context. *J-SHMC: Journal of English for Academic*, 13(1), 1–14.
- Iwata, A. (2026). Student perceptions of a speaking course enhanced by extensive reading within a self-determination theory framework. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688261457494>
- Jo'raxonova, M. (2025). Beyond the script: Assessing the impact of teacher enthusiasm on task completion and fluency in EFL classrooms. *International Scientific-Online Conference: Intellectual Education Technological Solutions and Innovative Digital Tools*, 4(48), 1–7.
- Krashen, S. D. (2004). *The power of reading: Insights from the research* (2nd ed.). Heinemann.
- Men, L., & Heng, K. (2026). Promoting extensive reading in EFL contexts: A review of benefits, challenges, and recommendations. *International Journal of Professional Development, Learners and Learning*, 8(1), Article e2608. <https://doi.org/10.30935/ijpdll/18053>
- Meniado, J. C. (2021). Integrating extensive reading in a college preparatory year program: Perceptions, challenges, and possibilities. *International Journal of English Language Education*, 6(2), 1–20. <https://doi.org/10.5296/ijele.v6i2.13624>
- Misikir, S., & Simegn, B. (2025). Incorporating extensive reading into EFL classroom lessons. *International Journal of Literacies*, 32(1), 115–130. <https://doi.org/10.18848/2327-0136/CGP/v32i01/115-130>
- Moazami Godarzi, F., Heidari Tabrizi, H., & Chalak, A. (2026). Impact of technology-enhanced personalized language learning on Iranian EFL learners' speaking skill: An introduction to Innovative Choice Learning Theory (ICLT). *International Journal of Practical and Pedagogical Issues in English Education*, 4(1), 1–17. <https://doi.org/10.22034/ijpie.2025.553906.1149>
- Moradi, P., & Zarei, A. A. (2025). The effects of synchronous computer-mediated interaction and face-to-face interaction on Iranian EFL learners' speaking and listening anxiety. *Journal of Modern Research in English Language Studies*, 12(4), 155–179. <https://doi.org/10.30479/jmrels.2025.21707.2489>
- Mostofi Shalmani, B., Pourkalhor, O., & Safdari, S. (2025). Tracing grammatical errors: A descriptive analysis of interlingual and intralingual interference in Iranian EFL learners. *Iranian Journal of Educational Research*, 4(3), 1–17. <https://doi.org/10.22034/ijer.2025.4.3.1>
- Putra, M. I. R., Irwan, D., Aunurrahman, A., Sulaiman, S., & Nurussaniah, N. (2026). Teaching with tech or text? A meta-analysis for evidence-based reading instruction. *International Journal of Educational Narrative*, 4(3), 826–848. <https://doi.org/10.70177/ijen.v4i2.3762>
- Rodrigo, V., Krashen, S., & Gibbons, B. (2004). The effectiveness of two comprehensible-input approaches to foreign language instruction. *System*, 32(1), 53–60.

- Sangers, N. L., van der Sande, L., Welie, C., Dobber, M., & van Steensel, R. (2025). Learning a language through reading: A meta-analysis of studies on the effects of extensive reading on second and foreign language learning. *Educational Psychology Review*, 37, Article 96. <https://doi.org/10.1007/s10648-025-10068-6>
- Sheu, S. P. H. (2003). Extensive reading with EFL readers at beginning level. *TESL Reporter*, 36(2), 8–26.
- Sun, X. (2022). Scaffolded extensive reading: A concept drawn from two EFL reading programmes. *Language Teaching Research*, 30(2). <https://doi.org/10.1177/13621688221141297>
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford University Press.
- Tabar, M. A., & Heydarnejad, T. (2026). Unearthing perfectionism and L2 speaking performance among Iranian EFL learners: The moderating role of cultural collectivism. *Journal of Psycholinguistic Research*, 55, Article 94. <https://doi.org/10.1007/s10936-026-10281-6>
- Tamersit, H., & Mouas, S. (2025). Critical classroom discourse analysis of EFL teachers and students at the Department of English Language and Literature – Batna 2 University. *Turkish Academic Research Review*, 10(1), 12–34. <https://doi.org/10.30622/tarr.1509484>
- Tlonaen, Z. A., Jaha, D. A. J., Ena, Z., & Dewi, R. F. (2025). Enhancing students' extensive reading learning outcomes: Integrating motivation, technology, and collaborative active learning. *Jurnal Pendidikan Progresif*, 15(3), 1989–2006. <https://doi.org/10.23960/jpp.v15i3.pp1989-2006>
- Xodabande, I., Shiri, S., & Zohrabi, M. (2025). Exploring the impacts of an AI-driven instructional intervention on Iranian EFL learners' pronunciation skill development. *Discover Education*, 4, Article 307. <https://doi.org/10.1007/s44217-025-00782-2>
- Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36(4), 661–672. <https://doi.org/10.1016/j.system.2008.04.003>
- Yang, A. (2001). Reading and the non-academic learner: A mystery solved. *System*, 29(4), 451–466. [https://doi.org/10.1016/S0346-251X\(01\)00038-0](https://doi.org/10.1016/S0346-251X(01)00038-0)
- Yazid, N. H. M., Sulaiman, N. A., & Hashim, H. (2025). Needs analysis of pre-university ESL students for the development of a web-based grammatical competence module. *The International Journal of Technologies in Learning*, 33(1), 1–23. <https://doi.org/10.18848/2327-0144/CGP/v33i01/1-23>
- Zohrabi, M., & Madineh, K. (2026). The supplementary use of English learning applications: Iranian EFL learners' attitudes, perceptions, and experiences. *Technology Assisted Language Education*, 4(1), 67–90. <https://doi.org/10.22126/tale.2026.13704.1176>



© 2026 by the authors. Licensee University of Zanjan, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license). (<https://creativecommons.org/licenses/by/4.0>).