Research in English Language Pedagogy (2025) 13(3): 130305

 DOI: 10.71673/relp.2025.1197426

Original Research

Using ChatGPT to Enhance Third-Person Narrative Grammar in Iranian EFL Learners

Yasaman Behravesh[1], *Hassan Alizadeh Mahmoud Alilo*[1*]
1Department of English, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Submission date: 25-01-2025 Acceptance date: 23-04-2025

Abstract

This research is a pioneering endeavor to understand how Iranian EFL students can benefit from utilizing ChatGPT to improve their acquisition of third-person narrative grammar. The study involved forty lower-intermediate female learners from Tabriz, Iran, with participants divided into two groups: an experimental group that used AI to enhance their learning, and a control group that received standard instructions. Participants' writing performance was thoroughly assessed through pre- and post-test assessments on grammatical concepts, including subject-verb agreement, use of pronouns, and tense. Throughout eight sessions, the experimental group engaged with ChatGPT for grammatical corrections and interactive exercises, while the control group followed conventional methods. The independent-sample t-tests revealed a statistically significant improvement in the experimental group's grammatical proficiency, as confirmed by two evaluators. Results indicated ChatGPT's potential in enhancing grammar accuracy and narrative coherence through adaptive feedback. Pedagogically, the study highlights the role of AI in supplementing EFL instruction, enhancing engagement, and improving grammar learning.

Keywords: ChatGPT, EFL Learner, Grammar, Narrative

[*] Corresponding Author's E-mail: Hassan1992alizadeh@gmail.com

1. Introduction

Emerging technology has paved the way for the complete replacement of the traditional learning environment with a new teaching methodology, where online instructors provide traditional teacher-led instruction. The outcome of this digital transformation of language education is an increase in the quality and availability of language education. As evidence indicates, using tech tools not only among students but also among teachers is a significant cause of learners' engagement and is one of the ways to individualize the learning process (Balcı, 2024). Through the adoption of digital tools in the EFL education scheme, the educational experience has advanced significantly, as these devices facilitate efficient language learning alongside the development of learner autonomy (Kucuk, 2024). This has been a breath of fresh air for teaching methods that provide appropriate student engagement and encourage students to relate concepts in the class. Such methods of delivery enable interaction with learners, making the educational process more productive and individualized.

A significant achievement in the educational field is the emergence of artificial intelligence tools, including generative pre-trained transformers (GPTs). In this group, ChatGPT has been in the spotlight due to its ability to mimic human behavior, produce grammatically correct content, and cater to the learning needs of an individual (Diasamidze & Tedoradze, 2024; Sallam, 2023). As a learning model, ChatGPT is a more general tool based on an advanced natural language processing (NLP) system with numerous real-world uses. Specifically, the proposed algorithm enhances essential language competencies, including reading, writing, and grammatical accuracy, demonstrating its efficacy in facilitating language acquisition (Kasneci et al., 2023). Its ability to think promptly with a customized approach is an invaluable means for evading the persistent hindrances that grammarians experience.

Although AI tools are increasingly applied in language education, the exploration of specific grammatical areas, such as third-person narrative writing, remains insufficiently addressed. This is in line with Amnuai et al. (2021), who note that achieving English proficiency is a fundamental imperative, compounded by challenging grammatical features, including subject-verb agreement, tense usage, and pronoun reference. However, other challenges that have been identified to affect grammar acquisition relate to language and cultural differences, and are even more challenging for EFL learners to overcome.

Given these persistent difficulties, innovative technological solutions, such as AI-based tools, have the potential to bridge the gaps in language learning. This study aims to address this gap by evaluating the efficacy of ChatGPT in improving grammatical precision and overall writing proficiency among learners of third-person narrative writing in Iran. The ability of an English language learner to use third-person narrative grammar effectively is a crucial aspect of advanced-level English proficiency. However, it is the same problem for English language learners worldwide(Shiu, 2011). In English, this kind of writing requires not only the competence to write narratives that are well-grounded and logical, but also interesting. The main grammatical issue, as pointed out by Caissie (1997), is subject-verb agreement, which is the most problematic for students, particularly those with a mother tongue that does not possess the same structures as relatively possessive pronouns, and inconsistent verb tenses. The research demonstrated that Korean EFL students make mistakes in using third-person singular verbs due to the absence of subject-verb agreement rules in Korean (Byon & Pyun, 2022). Additionally, Arab-speaking learners are often troubled by pronominal changes and verbal concord, which are closely tied to the Arabic structure of morphemes (Taj et al., 2017).

The problems faced by Iranian EFL learners in this area are exceedingly acute. In Persian, the predominant language in Iran, subject-verb agreement is not expressed in the same way as in English, thus causing frequent errors with the third-person singular form. It was pointed out by Rafiee and Abbasian-Naghneh (2021) in their study that Iranian people would mostly overlook the -s suffix in some verbs of the third-person singular. It is also important to note that these mistakes primarily stem from interference at the linguistic level, as well as the traditional approaches to teaching that Iranian education retains. The grammatical side of learning often involves rote-based learning accompanied by some grammar rules that are limited in their application to specific contexts (Rezaei et al., 2019). Thus, students might, for one thing, practically go through the process of properly learning grammar rules without actually gaining the experience.

There is another component of the stylistic and organizational features of the narratives, which complicates the pronunciation of the language. For instance, Ellis (2015) found that many EFL learners struggle to achieve accuracy in both the grammatical and the ideological aspects of writing. This is frequently connected with students' inadequate use of grammar in real-life situations, something that traditional teaching methods rarely

achieve. The consequence of the gap between students' theoretical proficiency and their practical performance highlights the importance of innovative strategies that address this issue.

To address the difficulties mentioned above, AI-based tools like ChatGPT offer a promising prospect for their successful resolution. ChatGPT helps students by providing them with immediate feedback and practice based on their mistakes. It is with the help of these AI tools that learning becomes interactive through their functions as opposed to static exercises, and all of this is achieved thanks to the benefits of these tools. The proposed model is also consistent with the constructivist theory of learning, as this theory suggests that involvement should be the primary focus of the learning process (Piaget, 1964; Vygotsky, 1978).

One important benefit of ChatGPT's capability in correcting grammatical errors is its customizability for speaking Farsi students who are at risk of lacking access to high-end educational technology. Given this, through its capacity to tailor its learning and training, it can be used to address potential learning issues that learners are likely to encounter based on the training context. For example, third-person singular verb agreement is a problem that cannot be well addressed using conventional traditional teaching aids. Additionally, by allowing students to test out ideas and receive immediate feedback, ChatGPT stimulates not only correct language but also the freedom of expression, which in turn addresses both the narrative writing′s dual needs: punctuation, subject-verb agreement, Parts of Speech, and imaginative provoking material.

While prior research has examined the general advantages of AI tools in EFL education, very few studies have investigated their impact on grammar-focused learning in specific contexts, such as third-person narrative writing. This study aims to fill this knowledge gap by investigating ChatGPT's effectiveness in enhancing both grammatical correctness and overall writing quality in Iranian EFL learners. It will also investigate learners' attitudes toward ChatGPT as a supplementary tool, measuring its practicality, accessibility, and potential for engagement and autonomy.

In this manner, this research contributes to a more comprehensive discussion on education in general and how AI can assist learners in Iran; it demonstrates how technological developments can be applied to meet specific linguistic and learning needs. Most crucially, it makes it clear that the utilization of AI tools has to be in harmony with

the general approach to teaching to avoid a mentality that deems smart technological solutions as sufficient substitutes for humans.

Thus, this paper provides evidence that language education can benefit from the implementation of AI-driven learning tools, as the application supplements learners' areas of difficulty, such as grammatical constructions, and refines their writing style. In this context, ChatGPT emerges as a practical application of relevant theoretical principles, serving as a valuable tool to bridge the gap between theoretical knowledge and practical implementation for both learners and educators. It also addresses limitations associated with conventional educational and training frameworks. The findings of this study will contribute to the continued advancement of the explored topic, specifically the application of AI in education, its potential to revolutionize EFL classroom practices, and the achievement of linguistic perfection.

2. Literature Review

The literature surrounding the use of ChatGPT to enhance third-person narrative grammar among Iranian EFL learners is at a burgeoning stage, characterized by increasing publication rates and growing interest within the academic community. While foundational studies exist, specific inquiries focusing on ChatGPT are still developing, highlighting both the opportunities and challenges for future research in this promising area. This suggests a fertile ground for in-depth exploration, encouraging further investigation and publication to bridge the existing gaps in the literature.

Research on personalized learning suggests that AI tools enable tailored educational experiences that cater to the individual needs of learners. ChatGPT's capability of providing customized feedback based on the learner's input is discussed as a pivotal point in enhancing grammatical competence. Critics argue that while personalization is beneficial, it may not adequately accommodate the diverse linguistic backgrounds of EFL learners, particularly in a culturally rich context such as Iran.

Related studies highlight the engagement and interactivity that AI offers, which may lead to increased motivation and enhanced personal learning experiences. Critical discussions highlight potential limitations, such as reliance on technology and the quality of AI-generated feedback. Technology has significantly impacted language teaching methodologies. Taj et al. (2017) explored technology-enhanced language learning,

demonstrating its effectiveness in improving grammatical accuracy among EFL learners. Their findings underscore the relevance of AI tools in facilitating real-time feedback, a key aspect of this study's examination of ChatGPT's role in enhancing third-person narrative grammar. Similarly, Rezaee et al. (2019) demonstrated that mobile-based dynamic assessments can enhance learners' oral accuracy, offering insights into how technology facilitates real-time feedback and personalized learning experiences.

Recent studies have explored ChatGPT's potential to transform EFL education. Kasneci et al. (2023) presented a case study highlighting ChatGPT's role in providing contextualized grammatical feedback, which is particularly relevant to this research. Their findings demonstrate that ChatGPT can effectively identify grammatical errors, offer rule-based explanations, and guide learners through corrective practices, thereby reinforcing grammatical accuracy in writing. This supports the current study by validating the use of ChatGPT as a tool for improving third-person narrative grammar, showing that AI-assisted feedback enhances learners' understanding and application of complex grammatical rules. These results align with Balcı's (2024) systematic review, showcasing its applicability across diverse linguistic contexts.

Moreover, Kucuk (2024) reported that ChatGPT-integrated grammar teaching significantly improved grammatical understanding and writing skills among EFL learners in Iraq, suggesting its adaptability to various educational settings. This study highlights the success of AI-assisted grammar instruction in a region with linguistic and educational similarities to Iran, reinforcing the feasibility of implementing ChatGPT in Iranian EFL classrooms. Similarly, Diasamidze and Tedoradze (2024) highlighted the efficacy of ChatGPT in enhancing the writing abilities of ESL students by providing real-time feedback and structured support throughout the writing process. Their work aligns with the present study by demonstrating how AI-driven assistance enhances learners' ability to produce coherent and grammatically correct written output. In the same vein, Nazeer et al. (2024) demonstrated that personalized conversational practice using ChatGPT improved learners' grammar accuracy and narrative coherence.

Research indicates that many EFL learners struggle with the subtleties of perspective, verb conjugation, and pronoun usage in third-person narratives. The debate within this theme centers on the effectiveness of instructional strategies and whether traditional methods can be supplemented or replaced by AI tools for improved grammar

acquisition. The complexity of grammar acquisition in EFL contexts has been widely documented. Nguyen and Tran (2023) identified challenges in applying grammatical rules in narrative writing, specifically in subject-verb agreement and verb tense consistency. These issues are echoed by Ellis (2015), who emphasized the importance of form-focused approaches to teaching grammar, highlighting the need for targeted interventions in both classroom and individual learning environments. Their arguments substantiate this study's rationale for incorporating ChatGPT as a supplementary tool to address persistent grammatical challenges in third-person narrative writing.

Caissie (1997) introduced cognitive prototypes for teaching verb phrase grammar to non-native speakers, a foundation that complements the AI-based approaches explored by recent studies. By integrating AI-driven feedback mechanisms, this study builds on Caissie's framework by demonstrating how ChatGPT can reinforce grammatical rules through personalized interactions. Pewngam et al. (2021) highlighted common structural issues in English writing among Thai students, including subject-verb agreement errors, which AI tools like ChatGPT can address through customized interventions. Their findings support the argument that automated feedback mechanisms are instrumental in improving learners' grammatical accuracy, aligning with the research goals.

Research on Iranian EFL learners has underscored the importance of technology readiness for successful integration. Rafiee and Abbasian-Naghneh (2021) developed a model to assess technology acceptance, indicating that learners' attitudes toward AI tools significantly influence their efficacy. These findings provide a crucial context for introducing ChatGPT into Iranian EFL classrooms to enhance narrative grammar. While ChatGPT holds significant promise, Sallam (2023) warned of potential limitations, including the risk of over-reliance on AI and ethical considerations in educational settings. These challenges necessitate careful implementation strategies to maximize the benefits of AI tools while mitigating their drawbacks.

Overall, the integration of ChatGPT into grammar instruction for Iranian EFL learners represents a promising fusion of technology and education, offering personalized, interactive, and engaging learning experiences. ChatGPT helps learners identify and correct grammatical errors in real time by providing instant feedback, which is particularly beneficial for grasping complex grammatical structures. Research highlights its effectiveness in reinforcing grammatical rules through examples and contextually relevant

scenarios, improving learners' writing skills and confidence. Beyond individual benefits, this approach democratizes access to high-quality educational resources, thereby fostering equitable learning opportunities, particularly in resource-constrained regions. However, future research must evaluate its long-term impact on grammar retention, compare traditional and AI-assisted methods, and explore broader demographic applications to ensure a robust understanding of its role in language education. To direct the investigation, this study formulates the following research question:

1. Does the use of ChatGPT as a grammar correction tool significantly impact the writing performance of EFL learners compared to those who do not use it?

3. Methodology

3.1. Design and Context of the Study

This study utilized a quasi-experimental approach with a pre-test and post-test control group design. The participants were divided into an experimental group, which received additional learning aids, and a control group, which received conventional instruction. The independent variable chosen in the present study was ChatGPT, while the dependent variable was the use of third-person narrative grammar in English. The study took place in a private language institution located in Tabriz, Iran.

3.2. Participants

Forty female EFL learners studying English were selected from the Pardisan Language Institute in Tabriz, Iran, from a total population of 120 learners during the 2023–2024 academic year. These participants, aged between 17 and 25, were native Turkish speakers with a minimum of two years of experience learning English at language institutes. A convenience sampling method was employed to select participants due to practical constraints such as accessibility and willingness to participate. This method ensured that the selected participants were representative of the general population of lower-intermediate EFL learners at the institute, as they shared similar educational backgrounds, language proficiency levels, and learning environments. To ensure homogeneity in English language proficiency, the Oxford Placement Test (OPT; Dave, 2004) was administered. Learners who scored between 30 and 39, classifying them as lower-intermediate, were chosen for the study. These 40 participants were then non-randomly assigned to either an

experimental group (n = 20) or a control group (n = 20), with both groups being instructed by the same teacher.

3.3. Instruments

3.3.1. Oxford Placement Test

The Oxford Placement Test (OPT) was administered to ensure the participants' English proficiency levels were comparable across both groups. This test was chosen because it is a widely recognized and standardized assessment tool that accurately measures language proficiency across different CEFR levels. One of the key features of the OPT is its ability to function as a homogenizing tool by assessing a range of linguistic competencies, including grammar, vocabulary, and reading comprehension. This test evaluates language proficiency using a six-point scale, categorizing scores as follows: Basic (A1: 0–17), Elementary (A2: 18–29), Lower Intermediate (B1: 30–39), Upper Intermediate (B2: 40–47), Advanced (C1: 48–54), and Very Advanced (C2: 54–60). Conducted at the beginning of the study, the test results were used to select only those participants who fell within the Lower Intermediate range (B1: 30–39).

3.3.2. Pre-Test and Post-Test Writing

The writing pre-test and post-test were critical instruments to measure participants' third-person narrative writing performance. The pre-test was administered one session prior to the treatment beginning. Students were asked to write a short third-person narrative paragraph (a story about a character). Similarly, after completing the eight-session treatment, the same test was re-administered as the post-test. The purpose was to assess the effect of the intervention on the students' writing skills. The pre- and post-tests were analyzed using a developed and comprehensive rubric. Third-person grammar identified by this rubric includes agreement, third-person subject pronouns, verb tense, subject-verb agreement, sentence construction, and grammatical correctness. The results ranged from 0 to 20, with higher values indicating a higher level of performance. Inter-ratter reliability was also used in the evaluation process to ensure validity. Two raters scored the tests to confirm the results. To maintain consistency, both raters engaged in a discussion to resolve any discrepancies in scoring. This approach ensured that differences in interpretation were addressed collaboratively, leading to a more reliable and standardized assessment process.

3.4. Data Collection Procedure

The following steps were taken to fulfill the objectives of this investigation and address the questions raised. First, one of the researchers, as a teacher, met with the institute's head to obtain authorization to conduct the study there. She introduced herself to the participants and encouraged them to participate in future research. She also informed them of their freedom to engage in this study freely. The course required students to attend twice a week for four weeks. Forty female EFL students in two classes participated in this study. A group of 15 learners, almost at the same level as the study's primary participants, were invited to pilot the tests. Adjustments were made based on their feedback. To maintain methodological rigor, all participants completed OPT at the outset. This test also helped standardize the sample because only participants whose language level corresponded to a certain norm were chosen. This step is taken to prevent interference of the baseline language skills in the study results. Then, they will be non-randomly assigned to an experimental group (n = 20) and a control group (n = 20).

Following the proficiency test, participants completed a writing pre-test. To determine the respective level of third-person narrative writing performance of the two groups, they were all given a pretest before the actual study began. In this phase, students were asked to write a short narrative paragraph in the third person (a story about a character). The tasks assigned to learners were completed without their knowledge of the actual research goals; therefore, they reflected their actual level of writing skills at the beginning of the process. Before initiating the treatment, the teacher explained the meaning and use of third-person narrative grammar, including subject-verb agreement and the avoidance of switching between first- and third-person pronouns and tenses, by reading examples of third-person narrative sentences and paragraphs. Afterward, the treatment was given to the experimental group for four weeks.

The experimental group in this study followed the AI-assisted learning procedures through ChatGPT. To achieve this, the teacher conducted a session before starting treatment, during which they familiarized students with ChatGPT and demonstrated how to interact with it, highlighting its potential applications, such as prompting and using ChatGPT to enhance the understanding and application of third-person narrative grammar for providing grammar feedback and generating examples. Throughout the sessions, the experimental group received immediate feedback from ChatGPT on grammatical errors,

with explanations and suggested corrections, allowing them to refine their writing iteratively. They followed the learning procedures outlined below. In the teaching phases of the research treatment, the activities were carefully designed to enhance students' understanding and application of third-person narrative writing through structured and interactive exercises. Initially, students were asked to think of a character or context, which was then used to prompt ChatGPT to write a brief third-person storyline. The output was reviewed by identifying how the correct grammar was used, noting errors/ambiguity, and discussing them. To achieve this, the features of third-person narratives, such as the appropriate use of pronouns and verb forms, were highlighted to the students within the generated text. Subsequently, students were given some third-person formatted prepared scripts filled with intentional grammatical mistakes. These texts were employed when working with ChatGPT on cooperative error annotation, in which the need for each correction has to be explained to enhance grammatical insight. Next, the students were given a task to rewrite the narratives with corrections as well as some further alterations that can improve the narrative coherence. Finally, students were tasked with rewriting the narratives with corrections and some further alterations that could improve the narrative coherence.

Conversely, the participants in the control group were instructed to study the same content using traditional methods. Following eight instructional sessions over four weeks, with two sessions per week, a writing post-test—similar to the initial assessment—was administered to evaluate the impact of ChatGPT on learners' writing performance. The authors themselves analyzed the participants' pre- and post-test writings, focusing on similar criteria to minimize subjectivity. The performance scores given by the raters were compared using the Pearson correlation coefficient to assess the inter-rater consistency. A high level of agreement indicated reliable scoring between raters. In cases of significant discrepancies, the two raters reviewed the evaluations and discussed them to reach a consensus. Finally, the collected data were entered into SPSS 27 for analysis to address the research question.

3.5. Data Analysis Procedure

The collected data were analyzed using SPSS 27. The OPT scores were first examined to confirm the initial homogeneity between the two groups. Subsequently, the means and

standard deviations of the pre-test and post-test scores were calculated. The Kolmogorov-Smirnov test was applied to assess the normality of the data. If the data followed a normal distribution, an independent samples t-test was conducted to evaluate the effect of ChatGPT on writing performance by comparing the baseline scores.

4. Results

The collected data were analyzed using SPSS version 27. An independent samples *t*-test was conducted to compare the average scores of the two groups on their written performance in both the pre-test and post-test.

4.1. Results of Pretest

A writing pretest was conducted to ensure the homogeneity of participants' writing performance prior to the main intervention. Two independent raters evaluated the pretest scores, and the results were analyzed for both groups. This step was necessary to ensure that any variation observed in post-test scores could be attributed to the intervention and not to any pre-existing variances in the measuring tool.

Table 1 presents the descriptive statistics for the pretest scores as rated by two raters. The data include the mean, standard error, and standard deviation of the mean for both groups.

Table 1

Descriptive Statistics of the Pretest Scores

Test	Group	N	Mean	Std. Deviation	Std. Error Mean
Pretest Rater 1	Experimental	20	11.375	.995	.222
	Control	20	11.462	.980	.219
Pretest Rater 2	Experimental	20	11.362	1.030	.230
	Control	20	11.350	.984	.220

Based on the findings presented in Table 1, the pretest results, as evaluated by two independent raters, demonstrate a high level of consistency and indicate homogeneity among the participants across the control and experimental groups. According to the first rater, the mean scores for the experimental group (M = 11.375, SD = 0.995) and the control group (M = 11.462, SD = 0.980) are nearly identical. The standard error of the mean (SEM) values for both groups (±0.222 for the experimental group and ±0.219 for the control group) further confirms the small variation in the sample scores. Similarly, the

scores rated by the second rater show minimal difference. The experimental group achieved an average score of 11.362 (*SD* = 1.030), while the control group achieved a mean of 11.350 (*SD* = 0.984). The SEM values (±0.230 for the experimental group and ±0.220 for the control group) indicate a narrow range of standard error across both groups.

The close similarity in means across both raters supports the reliability of the pretest evaluation process. Moreover, the near-zero standard deviations presented for both groups, as well as the further low standard error of the means, suggest that, presumably, both groups performed at equal levels before the intervention. Such matches are a measure that ensures experimental and control groups start the study with few initial differences to influence the results of the primary research. The pretest results evaluated by two independent raters confirm the uniformity of the experimental and control groups. The minimal differences in mean scores and the consistent standard deviations highlight the reliability and objectivity of the raters' evaluations. Consequently, subsequent views of post-test outcomes are more likely to be attributed to the independent impact of the intervention rather than participants' initial pretense differences in abilities.

To assess the consistency between the two raters, the Pearson correlation coefficient was utilized to determine inter-rater reliability. The results of this analysis, specifically for the pretest scores, are presented in Table 2, which outlines the inter-rater reliability metrics.

Table 2

Inter-Rater Correlation for the Pre-test Scores of Both Groups

		Pretest Rater 1	Pretest Rater 2
Pretest Rater 1	Pearson Correlation	1	.945**
	Sig. (2-tailed)		.000
	N	40	40
Pretest Rater 2	Pearson Correlation	.945**	1
	Sig. (2-tailed)	.000	
	N	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

Table 2 presents the inter-rater reliability of the pretest scores, as evaluated by two raters, using Pearson correlation coefficients. The analysis reveals a strong level of agreement among the raters, as indicated by a Pearson correlation coefficient of 0.945. This strong correlation indicates that the pretest scores assigned by both raters are consistent and reliable. The statistical significance ($p < 0.01$) further supports the

robustness of the correlation. Given this high inter-rater reliability, both raters evaluated the participants' pretest performance with a comparable level of accuracy and objectivity.

Parametric statistical tests, such as the independent sample t-test, require a normality test to ensure that the data do not significantly deviate from a normal distribution. The normality of the pretest scores for both raters was evaluated using the Kolmogorov-Smirnov and Shapiro-Wilk tests to assess compliance with the normality assumption. The Kolmogorov-Smirnov test was chosen because it is effective for larger sample sizes and assesses whether the data significantly deviates from a normal distribution by comparing it to an ideal normal cumulative distribution. In contrast, the Shapiro-Wilk test is particularly useful for smaller samples due to its sensitivity in detecting deviations from normality. Using both tests provided a more comprehensive assessment of the normality assumption in this study. These results are provided in Table 3.

Table 3

Tests of Normality for Pretest Scores

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pretest Rater 1	.141	40	.044	.951	40	.080
Pretest Rater 2	.191	40	.001	.954	40	.107

a. Lilliefors Significance Correction

Table 3 presents the results of two normality tests. For the first rater, the Kolmogorov-Smirnov test yielded a statistic of $D(40) = 0.141$, $p = 0.044$, indicating statistical significance ($p < 0.05$), while the Shapiro-Wilk test produced a statistic of $W = 0.951$, $p = 0.080$, which was not significant ($p > 0.05$). Similarly, for the second rater, the Kolmogorov-Smirnov test yielded $D(40) = 0.191$, $p = 0.001$, indicating significance ($p < 0.05$), whereas the Shapiro-Wilk test reported $W = 0.954$, $p = 0.107$, which was not significant ($p > 0.05$). When results from the two tests conflict, the Shapiro-Wilk test is generally preferred due to its higher sensitivity, especially for smaller sample sizes, such as this study ($n = 40$). Based on the Shapiro-Wilk outcomes, the pretest scores for both raters can be considered normally distributed. Hence, based on the results of the Shapiro-Wilk test, it could be concluded that the pretest scores assessed by both raters conform to the normality assumption as required for the application of parametric tests in the subsequent data analysis.

An independent-samples *t*-test was performed to compare the pretest scores of the groups for both raters. Prior to the *t*-test, Levene's test was conducted to assess the equality of variance between the two groups. These analyses are outlined in Table 4.

Table 4

Results of the Independent-Sample t-test on both Groups' Pretest Scores

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pretest Rater 1	Equal variances assumed	.002	.964	-.280	38	.781	-.087	.312	-.720	.545
	Equal variances not assumed.			-.280	37.992	.781	-.087	.312	-.720	.545
Pretest Rater 2	Equal variances assumed	.076	.784	.039	38	.969	.012	.318	-.632	.657
	Equal variances not assumed.			.039	37.921	.969	.012	.318	-.632	.657

As presented in Table 4, Levene's test for the first rater's pretest scores indicated no significant difference in variances between the groups ($F = 0.002$, $p = 0.964$), allowing the assumption of equal variances. The t-test results revealed no significant difference between the experimental group ($M = 11.375$, $SD = 0.995$) and the control group ($M = 11.462$, $SD = 0.980$), $t(38) = -0.280$, $p = 0.781$. The mean difference was -0.087, with a 95% confidence interval ranging from -0.720 to 0.545. Similarly, for the second rater's pretest scores, Levene's test confirmed equal variances ($F = 0.076$, $p = 0.784$). The t-test results indicated no significant difference between the Experimental group ($M = 11.3625$, $SD = 1.03070$) and the Control group ($M = 11.350$, $SD = 0.984$), $t(38) = 0.039$, $p = 0.969$. The mean difference was 0.012, with a 95% confidence interval ranging from -0.632 to 0.657. Overall, the independent-sample t-test results for both raters show no significant differences between the pretest scores of the Experimental and Control groups. These findings confirm that the two groups were homogeneous in their performance before the intervention, as both raters provided consistent evaluations of the scores.

4.2. Results of Post-test

To address the research question regarding the significant difference in writing performance between EFL learners who use ChatGPT as a grammar correction tool and those who do not, an independent-sample t-test was conducted. The descriptive statistics for the post-test results are presented in Table 5.

Table 5
Descriptive Statistics of the Post-test Scores

Test	Group	N	Mean	Std. Deviation	Std. Error Mean
Posttest Rater 1	Experimental	20	18.200	1.260	.281
	Control	20	14.600	1.220	.272
Posttest Rater 2	Experimental	20	18.250	1.164	.260
	Control	20	14.637	1.270	.284

Table 5 presents the descriptive statistics for the post-test scores, as assessed by two raters, for both groups. For the first rater's post-test scores, the experimental group achieved a mean score of 18.20 (*SD* = 1.26, *SEM* = 0.28), while the control group had a mean score of 14.60 (*SD* = 1.22, *SEM* = 0.27). Similarly, for the second rater's post-test scores, the experimental group obtained a mean score of 18.250 (*SD* = 1.164, *SEM* = 0.260), whereas the control group had a mean score of 14.637 (*SD* = 1.270, *SEM* = 0.284). These findings indicate that the experimental group consistently outperformed the control group in the post-test assessments, as evidenced by the higher mean scores from both raters.

Table 6 presents the inter-rater reliability for the post-test scores of both groups, as assessed using the Pearson correlation. The results indicate a strong level of agreement between the two raters.

Table 6
Inter-Rater Correlation for the Post-test Scores of Both Groups

		Posttest Rater 1	Posttest Rater 2
Posttest Rater 1	Pearson Correlation	1	.993**
	Sig. (2-tailed)		.000
	N	40	40
Posttest Rater 2	Pearson Correlation	.993**	1
	Sig. (2-tailed)	.000	
	N	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

As Table 6 represents, for both raters' scores on the post-test, the Pearson correlation coefficient was r = 0.993, with a p-value of 0.000. This demonstrates a strong, positive, and statistically significant correlation ($p < 0.01$) between the scores assigned by the two raters.

The normality of the post-test scores for both raters was evaluated using the Kolmogorov-Smirnov and Shapiro-Wilk tests to assess whether the data met the assumption of normality (see Table 7).

Table 7

Tests of Normality for Posttest Scores

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Posttest Rater 1	.097	40	.200*	.964	40	.226
Posttest Rater 2	.079	40	.200*	.961	40	.187

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As Table 7 demonstrates, for the post-test scores of the first rater, the Kolmogorov-Smirnov test resulted in a statistic of 0.097 with a significance value of p = 0.200. The Shapiro-Wilk test yielded a statistic of 0.964 with a p-value of 0.226. Similarly, for the post-test scores of the second rater, the Kolmogorov-Smirnov test yielded a statistic of 0.079 with a significance value of p = 0.200, while the Shapiro-Wilk test produced a statistic of 0.961 with a significance value of p = 0.187. As the significance values ($p > 0.05$) for both raters across both tests exceed the threshold, the results indicate that the post-test scores for both raters meet the normality assumption. Therefore, the data is suitable for parametric statistical analysis.

Table 8 presents the results of the independent-sample t-test conducted to compare the post-test scores of groups, as evaluated by two raters. Prior to performing the *t*-test, Levene's Test for Equality of Variances was conducted to confirm that the assumption of homogeneity of variances was satisfied.

Table 8

Results of Independent-Sample t-test on both Groups' Post-test Scores

		Levene's Test for Equality of Variances		t-test for Equality of Means					
								95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Posttest Rater 1	Equal variances assumed	.018	.893	9.175	38	.000	3.600	.392	2.805	4.394
	Equal variances not assumed			9.175	37.960	.000	3.600	.392	2.805	4.394
Posttest Rater 2	Equal variances assumed	.390	.536	9.374	38	.000	3.612	.385	2.832	4.392
	Equal variances not assumed			9.374	37.712	.000	3.612	.385	2.832	4.392

As Table 8 illuminates, for the post-test scores of the first rater, Levene's test resulted in an F-value of 0.018 with a significance level of 0.893, which is greater than 0.05, confirming that the variances between the two groups are equal and satisfy the assumption of homogeneity. The t-test revealed a *t*-value of 9.175 with 38 degrees of freedom and a significance level (p-value) of 0.000, indicating a statistically significant difference between the experimental and control groups. The mean difference between the two groups was 3.600, with a standard error of 0.392, and the 95% confidence interval for the mean difference ranged from 2.805 to 4.394. For the post-test scores of the second rater, Levene's test also confirmed the equality of variances, with an F-value of 0.390 and a significance level of 0.536. This result satisfied the assumption of homogeneity for the *t*-test. The *t*-test for Rater 2 showed a t-value of 9.374 with 38 degrees of freedom and a significance level of 0.000, further supporting a statistically significant difference between the two groups. The mean difference was 3.61250, with a standard error of 0.385, and the 95% confidence interval ranged from 2.832 to 4.392, reaffirming the significant difference in post-test scores.

In conclusion, the results of the independent sample t-test for both raters indicate that the experimental group performed significantly better than the control group. Both raters reported p-values of 0.000, indicating strong statistical evidence of a significant difference in post-test scores between the two groups. The mean differences of 3.600 and 3.612 for the first and the second rater, respectively, highlight the superior performance of the experimental group. The consistency of the results across both raters reinforces the reliability of these findings and suggests that the experimental intervention was effective in improving post-test scores.

5. Discussion

This research investigated the potential of ChatGPT as a complementary teaching method for third-person narrative grammar in Iranian EFL learner classrooms, comparing it with traditional instruction. Results indicated a marked increase in the linguistic correctness of the group under study. The thorough behaviors achieved by the independent raters were a mark of that enhancement. This finding suggests the potential of AI-powered teaching tools to help solve the intricacies of grammar in language learning communities.

The experimental group utilized ChatGPT and achieved significant advancements in specific grammar topics, including third-person writing. These involved using the correct verb, referring to the noun with the correct pronoun, and ensuring the correct verb tense. The findings of this study align with existing research that provides evidence for the effectiveness of computer-assisted grammar therapy in improving young learners' correct use of conjunctivitis and present and past forms (Kasneci et al., 2023; Sallam, 2023). The demonstration of ChatGPT as an effective means of such guided practice was possible via the fact that this tool is capable of immediate as well as personalized feedback; thus, students have a chance not only to spot, analyze, and resolve but also to get through authentic writing tasks, and thus can experience how language learning through writing is done. One of the common characteristics of such setups is the noticeable absence/undeveloped state of interactivity in traditional classroom environments, which are usually characterized by the fact that personalized support is scarce due to resource limitations and time constraints (Nguyen & Tran, 2023).

Contrarily, and in stark contrast to traditional grammar and rule instructions, which by their very nature are generally less dynamic, the interaction in ChatGPT is active and

adaptive. Traditional classes often rely heavily on plain note memorization and decontextualized grammar drills, which limit their ability to apply such rules in real social settings. ChatGPT is an approach that enables learners to feel more involved and learn actively within a contextualized environment, which is of great importance for constructivist learning. These theories stem from the fact that constructivist learners can learn effectively when they are involved, rather than being mere recipients of information (Piaget, 1964; Vygotsky, 1978).

ChatGPT contextualizes and makes these theoretical understandings more iterative, leading to practical applications. It therefore encourages learners to try using the language, receive feedback, and refine their understanding in an iterative process that enhances not only grammatical accuracy but also deeper retention of the language rules.

These results are supported by other research demonstrating the effectiveness of generative AI in language learning. In other instances, Diasamidze and Tedoradze (2024) present a case for enhanced grammatical accuracy in EFL learners using AI-enhanced systems, while Kucuk (2024) highlights the benefits of a technology-enhanced learning environment within a broader framework of developing linguistic competence and enhancing learners' independence. This work's originality lies in its focus on third-person narrative grammar, which has hardly received attention from Iranian EFL learners.

Linguistic challenges that Persian speakers encounter while learning English grammar are well-documented. Grammatical structures not found in Persian, such as the third-person singular verb agreement identified by Rafiee and Abbasian-Naghneh (2021), often constitute the source of persistent errors among learners. The improvements observed in this study suggest that ChatGPT helps bridge this linguistic gap. It enables learners to internalize complex grammatical rules, which may otherwise elude them, through the simple practice of repetition in context.

This paper highlights, from a pedagogical perspective, the transformative potential of integrating AI tools, such as ChatGPT, into EFL curricula. These tools complement traditional teaching methods with their personalized and adaptive nature, thus offering learners further opportunities for targeted practice and feedback. Research by Kasneci et al. (2023) and Kucuk (2024) supports this notion, demonstrating that AI-driven language learning tools can significantly improve grammatical accuracy and learner engagement. ChatGPT has the potential to be a useful tool for educators in creating interactive grammar

exercises, designing individualized learning plans, and addressing specific learner needs. According to Ellis (2015), effective grammar instruction requires both explicit rule teaching and opportunities for meaningful practice. ChatGPT aligns with this approach by providing real-time feedback, allowing learners to internalize grammatical structures through repeated exposure and correction. Additionally, Taj et al. (2017) highlighted the role of technology-enhanced learning in improving grammar acquisition, reinforcing the argument that AI tools like ChatGPT can serve as valuable supplementary resources in EFL classrooms.

Iranian EFL learners, in particular, see the bot as a low-cost and scalable solution to their issues with traditional grammar teaching. Due to the lack of learning resources, most students in this situation do not get a personalized learning approach (Rezaei et al., 2019). By offering essentially live and interactive education, students also have the opportunity to study through ChatGPT, which, in turn, will broaden the spectrum of learners from diverse backgrounds.

Furthermore, this research endorses the general introduction of AI tools into communicative language teaching (CLT) methods. By saturating grammar instruction in meaningful and interactive contexts, ChatGPT enables learners to develop both grammatical and communicative competencies, a vital objective of contemporary language education.

The results of this research are positive; however, several limitations need to be recognized. The sample size was relatively small (n=40), and the population was homogeneous, consisting only of female learners from Iran who had Persian as their first language. Hence, the study's findings can hardly be generalized to broader populations. Subsequent studies should expand their scope to not only increase sample size but also make them more diverse, in order to establish whether the same results can be achieved in different demographic and linguistic contexts. Another limitation is that the intervention only lasted four weeks. Although significant improvements were observed during this timeframe, the long-term implications of AI-assisted learning on grammar retention remain unknown. Longitudinal studies are necessary to determine whether the gains achieved with ChatGPT are sustained over time and how they impact learners' overall language proficiency.

Future research could explore the application of ChatGPT to other areas of grammar or writing, such as descriptive, argumentative, or expository texts. Furthermore, qualitative studies on learners' perceptions of AI tools could provide more insight into the motivational and engagement variables that determine their success. Understanding learners' attitudes and experiences could help create more user-friendly and pedagogically successful AI interventions.

Finally, investigations comparing the efficacy of ChatGPT to other AI tools or traditional instructional techniques could provide insight into optimal practices for incorporating technology into language education. Such research may investigate the optimal duration and frequency of AI-assisted interventions, the types of feedback that learners find most useful, and the role of instructor assistance in amplifying the impact of AI tools.

Overall, this study provides persuasive evidence that ChatGPT can be an effective additional tool for teaching third-person narrative grammar to Iranian EFL learners. ChatGPT addresses specific grammar issues that traditional approaches often fail to resolve, providing instant, personalized feedback and encouraging active engagement. These findings contribute to the growing body of research on AI in language education, underscoring its potential to improve learning outcomes and bridge linguistic and pedagogical gaps.

6. Conclusion

This study investigated the efficacy of ChatGPT in improving third-person narrative grammar among Iranian EFL learners, specifically focusing on subject-verb agreement, tense usage, and pronoun reference. The findings affirm that ChatGPT significantly enhances learners’ grammatical accuracy and narrative writing skills compared to traditional instruction methods. The experimental group demonstrated notable improvements in the post-test results, reinforcing the hypothesis that AI-driven tools can connect theoretical knowledge with practical grammar use in EFL contexts.

One of the most compelling outcomes of this research is the personalized and interactive learning environment that ChatGPT fosters. Unlike conventional strategies, often rooted in repetitive memorization and decontextualized grammar drills, ChatGPT provides real-time feedback, contextualized examples, and iterative opportunities for

correction. These features not only improve grammatical accuracy but also build learners' confidence and autonomy in narrative writing. The integration of ChatGPT aligns with constructivist theories of learning, which emphasize active involvement and contextualized understanding as key to effective language acquisition.

The research also highlights the specific linguistic difficulties faced by Iranian EFL learners, such as difficulties with third-person singular verb agreement, which stem from structural differences between Persian and English. ChatGPT's capacity to provide customized feedback tailored to these specific challenges makes it a valuable tool in addressing these persistent errors. By offering targeted practice and immediate corrections, ChatGPT helps learners internalize complex grammatical rules, thus enhancing their overall writing proficiency.

Moreover, this study enhances the evolving discourse on integrating AI tools in language education. It highlights the potential of ChatGPT to democratize access to quality language learning resources, particularly in resource-constrained environments. For Iranian learners, ChatGPT offers a scalable and cost-effective solution to overcome these limitations, where traditional grammar instruction often lacks interactivity and personalization.

In conclusion, this research underscores the transformative potential of AI-driven tools like ChatGPT in enhancing EFL grammar instruction. By integrating these technologies into the curriculum, educators can create more engaging, effective, and learner-centered experiences that cater to individual needs. As the role of AI in education continues to evolve, tools like ChatGPT promise to revolutionize language learning, making it more accessible, interactive, and tailored to the diverse needs of learners in an increasingly interconnected world.

References

Amnuai, W., Pewngam, C., Nawatmongkolkorn, P., & Pimpa, K. (2021). Analyses of structural organization and subject-verb agreement of English paragraphs written by Thai accounting students. *Rangsit Journal of Educational Studies*, *8*(2), 15-29.

Balcı, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies 3*(1), 48-63.

Byon, A. S., & Pyun, D. O. (2022). *The Routledge handbook of Korean as a second language*. Routledge.

Caissie, R. A. (1997). *English verb phrase grammar prototypes for speakers of other languages: A cognitive approach to facilitate second language English composition*. University of Washington.

Dave, A. (2004). *Oxford placement test 2: Test pack*. Oxford University Press.

Diasamidze, L. and T. Tedoradze (2024). Enhancing ESL students' writing skills through the natural language processing model ChatGPT. *The Eurasia Proceedings of Educational and Social Sciences, 35*(2), 230-238.

Ellis, R. (2015). Form-focused approaches to learning, teaching, and researching grammar. In *Teaching and Learning English Grammar* (pp. 194-213). Routledge.

Kasneci, E., et al. (2023). ChatGPT as a tool for language education: A case study. *AI in Education Research. 4*(6), 58-69.

Kucuk, T. (2024). "ChatGPT Integrated Grammar Teaching and Learning in EFL Classes: A Study on Tishk International University Students in Erbil, Iraq." *Arab World English Journal (AWEJ) Special Issue on ChatGP*, 100-111. https://dx.doi.org/10.24093/awej/ChatGPT.6

Nazeer, I., Yasmin, S., & Khan, N. M. (2024). English language acquisition through ChatGPT: A study on personalized conversational practice. *Journal of Applied Linguistics and TESOL (JALT), 7*(4), 1076-1091.

Nguyen, T., & Tran, H. (2023). The challenges of grammar application in narrative writing. *EFL Pedagogy Quarterly. 5*(9), 108-124.

Piaget, J. (1964), *The construction of reality in the Child*. New York.

Rafiee, M., & Abbasian-Naghneh, S. (2021). E-learning: developing a model to assess the acceptance and readiness of technology among language learners. *Computer Assisted Language Learning*, *34*(5-6), 730-750.

Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2019). The impact of mobile-based dynamic assessment on improving EFL oral accuracy. *Education and Information Technologies, 24*(5), 3091-3105

Sallam, M. (2023). The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *MedRxiv*, 2(2), 23-37. https://doi.org/10.1101/2023.02.19.23286155

Shiu, L. J. (2011). *EFL learners' perceptions of grammatical difficulty in relation to second language proficiency, performance, and knowledge*. University of Toronto.

Taj, I. H., Ali, F., Sipra, M. A., & Ahmad, W. (2017). Effect of technology enhanced language learning on grammar acquisition of EFL learners. *International Journal of Applied Linguistics and English Literature, 6*(3), 262-272.

Vygotsky, L. (1978), *Mind in society: The development of higher psychological processes.* Harvard University Press, Cambridge, MA.