Vol. 3 No. 4 (2025)

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895
https://llrjournal.com/index.php/11

A Quasi-Experimental Study on the Effect of ChatGPT in ESL Grammar Instruction

Shagufta Kanwal
Hamdard University Karachi
Email: kanwalmehar111@gmail.com

Hafiz Imran Nawaz
Senior Lecturer, Hamdard University Karachi
Email: imran.nawaz@hamdard.edu.pk

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Abstract

The study investigates the influence of ChatGPT on the English grammar proficiency of students from the Federal Board of Intermediate and Secondary Education (FBISE) in Pakistan. A quasi-experimental research design consisting of two intact groups, a control group and an experimental group, each with 30 students, was employed. The experimental group was taught with the help of ChatGPT and the control group was taught with conventional methods. Over the ten-week intervention, the focus was on three main aspects of grammar: the types of sentences, phrases and clauses, and narrations. A pre-test, post-test and was given to the students. The results of the study were analysed using parametric tests with the independent and paired samples. Results of the study indicated that while both groups showed considerable improvement, the experimental group showed even greater improvement as compare to the control group. The study contends that the use of ChatGPT in teaching grammar results in enhanced grammar proficiency of the students. Results of the study showed that there is an existing and potential impact of using AI as a teaching tool in Pakistani classrooms and that study results can be used as a guide by language teachers, curriculum designers, and policymakers to meet the emerging and existing requirements of teaching and learning in Pakistan.

Keywords: ChatGPT, ESL, Artificial Intelligence, English Language Learning, English Grammar Skills.

Introduction:

In a globalised economy, English proficiency acts as a prerequisite for educational achievement and career opportunities (Pajares & Valiente, 1996, p. 268). The ability to communicate in English competently and efficiently and to utilise a variety of English-language communication strategies is indispensable for developing learning behaviours at postsecondary and advanced levels in the global economy (Rao, 2019; Dukhan & Joghan, 2021). Acquiring English grammar is an essential academic requirement primarily for students in Pakistan, and most specifically in the Federal Board of Intermediate and Secondary Education (FBISE) matriculation programmes (Ali, 2019; Heydari and Bagheri, 2015). Despite the importance of learning English grammar, the approach used in its teaching is still largely artisanal, with an excessive emphasis on the teaching of grammar (Songsiengchai et al., 20203). Such a rigid educational model fails to meet the students' needs, especially in an English as a Second Language (ESL) context.

In Pakistan, particularly for students in the Federal Board of Intermediate and Secondary Education (FBISE) matriculation programmes, the ability to learn English grammar is a requirement in the curriculum as well as a fundamental skill for continued education and global employment (Ali et al., 2020). However, even though it remains important, the instruction of English grammar tends to be traditional as it is anchored in rote learning and rules-based grammar instruction (Songsiengchai et al., 2023). Such a system hardly addresses the evolving demands of the students, particularly in an ESL context.

In Pakistan, the use of the English language differs in relation to social class, location,

and the level of educational institution. Learning English starts at early grades in school, yet many students who complete standard 10 (matriculation level) have serious issues when it comes to English grammar and writing (Sajid and Siddiqui 2015). It can be attributed to many factors such as lack of effective teaching, absence of relevant educational materials, inadequate teaching resources, and teachers dominating the classroom (i.e. authoritarian) and failing to accommodate the different learning styles of their students. The role of English as a language in the education of Pakistan is both critical and quite dysfunctional (Haider and Fang 2021) Under the Federal Board (FBISE), students must have the ability to use English at an acceptable level as to grammar, comprehension and writing skills when they complete their education at 10th standard (matriculation); however, the level of teaching, and of students themselves in English, is, at the best of times, very poor (Baig et al 2019).

ESL Pakistan Learners' Grammar Skills

As far as learned a second language and acquired grammar are concerned, for Pakistani ESL students, it is a prerequisite and a challenge. As a case in point, for communicative competence, ESL students need to be able to produce functionally and grammatically correct sentences. Unfortunately, knowing a language is not enough to fully understand that language. In Pakistan, teaching of grammar has remained stuck to rote learning of rules without contextual application, as pointed out by (Bano, 2025). Hence, Richards (2006) posits that ESL learners have a declarative knowledge of rules of grammar, while lacking it in procedural knowledge to be able to use them in written and spoken discourse.

Students often find grammar abstract and disconnected from real language use. For instance, although they may know the definition of a compound-complex sentence, they rarely use it effectively in their writing. Common grammatical errors among matric students include incorrect tense usage, subject-verb agreement mistakes, fragmented sentences, run-ons, and misuse of articles and prepositions (Khosa, 2021). Such errors persist because students receive insufficient feedback and do not engage in active, meaningful grammar practice. ChatGPT presents new opportunities to enhance grammar teaching to ESL leaners.

The Role of ChatGPT in Improving Students' Grammar Skills

ChatGPT is able to understand and model human-like text, therefore becoming an effective language learning tool. It can reinforce grammar learning by giving feedback in near real-time, providing contextual explanations, and allowing custom practice which are critical in enhancing learner autonomy and language proficiency (Prather et al., 2024).

The present research aims at gaining insight into the role of ChatGPT in enhancing the level of English grammar in matriculation students at the Federal Board of Pakistan in reference to the three components of grammar; use of different types of sentences, proper use of tenses, usage of phrases and clauses, as well as comprehension of narratives. Having provided a detailed analysis of the study, this research is capable of exploring the ways of how ChatGPT will be able to correct what is wrong with the existing method of teaching grammar and improve grammatical accuracy of the students in those very areas.

Objectives of the Study

The aims of the objectives are:

To evaluate the overall impact of AI-assisted learning through ChatGPT on students' grammar proficiency

To examine the impact of ChatGPT on students' ability to identify and construct various types of sentences

To analyze the effectiveness of ChatGPT in the usage of phrases and clauses

To assess the role of ChatGPT in improving students' proficiency of narrations

Significance of the Study

There are different stakeholders within the educational community, such as educators, administrators, and researchers, who will each benefit from the findings of this study. Primarily, the study's examination of ChatGPT's effectiveness at improving grammatical accuracy contributes to the discourse on the practicality of AI tools for teaching and learning language skills. Furthermore, instructors can implement the results of this research to understand how ChatGPT should be used to teach grammar. Supposing that it is effective, ChatGPT can be incorporated into classroom instructions in order to deliver individual and interactive learning, and solve the weaknesses of traditional learning. The societal level would prove useful in steering the psyche of policymakers to understand the benefits of integrating AI technologies into the national curriculum. It states that it is necessary to invest in digital infrastructure and educate teachers to make all students equal users of AI tools. The paper emphasizes the need to embrace new technology such as AI in solving the plight of matriculation students in Pakistan. It opens the way to further research on the role of AI in other areas of language learning and instruction.

Literature Review

Traditinal Language Teaching Methods

Conventional methods have been the foundation of teaching curriculums in most of the educational systems in the world, including Pakistan. GTM has been traditionally the most popular one in secondary schools of Pakistan. It focuses more on learning grammatical rules, memorising vocabulary, and translating material between English and the native language of the learner usually Urdu. It encourages analytical skills through emphasis on syntax and morphology and is compatible with exam-driven teaching in which the accuracy of grammar can be rewarded (Bhatti & Mukhtar, 2017). Its key drawbacks are that it does not focus on the listening and speaking skills that are part and parcel of actual communication. In addition, learners acquire inert knowledge without being able to effectively communicate using the language in real life situations (Tabassum, 2020). Next, the Direct method can hardly be used universally due to the lack of trained teachers, teaching resources, and overpopulated classrooms. Although the pedagogical merits of Direct Method are hard to deny, it is hardly ever applied outside elite private institutions due to infrastructural and logistical challenges (Mohammad et al., 2018). When it comes to The Audio-Lingual Method, it is strong in its ability to help learners to improve their listening and pronunciation, putting words together because of repetition and drill dialogues. Regardless, this method does not intellectually appeal to learners, as it is rote memorisation, rather than understanding. However, CLT is considered a forward and

effective approach to achieving communicative competence. It is learner-centred and falls in line with contemporary SLA theories which posit that meaningful interaction is vital for language learning. Still, the effectiveness of CLT is largely dependent on the state of the classroom, the teacher's proficiency, and the system of evaluation. In Pakistan, the implementation of this is complicated due to the large number of students in a classroom, the inflexible nature of the syllabus, lack of materials, and a grammar-centred approach to communication in the evaluation system (Wasti, 2016).

The examination based system of education has led to continuation of traditional methods. To a large extent, this kind of standardised assessment used in Pakistan focuses not on language fluency or language use in a real situation, but on grammatical accuracy, translation tasks, and written rote. This has led to the teaching practices which emphasize rote study as opposed to communicative competence (Shamim & Tribble, 2005). Second, many of the public-sector schools lack facilities which make it hard to adopt modern teaching methodology. Schools are poorly staffed, insufficiently financed and often deprived of technological resources like multimedia equipment or language labs, and have too scanty access to updated textbooks or teacher training programs. According to Hussain et al. (2020), these conditions do not allow teachers to adopt activity-based or student-centered learning activities such as Communicative Language Teaching or Total Physical Response that necessitate the availability of resources and flexibility of learning environments. In this regard, Artificial Intelligence has greatly changed schooling, especially in learning languages.

AI in Language Learning

AI greatly influences the way students learn languages, also the way teachers present content and, finally, the manner in which institutions organize their educational activities (Pedro et al., 2019). In outstanding similarity with that, the U.S. Department of Education published a 2023 report, “Artificial intelligence and the future of teaching and learning: insights and recommendations”, in which the collective effort to embrace AI was noted as critical. Through the responses of more than 700 stakeholders, the report asserted that AI has the promise of increasing scalability, cost reduction, and aligning with educational priorities, as long as its risks are addressed (Yang et al., 2024).

The use of AI to improve second language learning results is another body of research whose quantity is growing. The AI applications in language learning harness cutting-edge algorithms, making learning individualized, exciting, and straightforward. Roberts and Jones (2019) emphasized that AI-based systems promote individualized experiences in learning as they are aligned to accommodate an individual based on his or her needs and levels of proficiency. Through an immersion in human-like conversations and presenting personalized answers, ChatGPT has created new opportunities to improve the learning experiences (K et al., 2024).

ChatGPT for Improving Grammar Skills in Writing

ChatGPT, being an AI-based tool has proven beneficial in improving writing skills among the learners. In one combined-methods intervention, ChatGPT was found effective for academic writing skills in being a reliable use of feedback tool, particularly in large courses (Mahapatra, 2024). Students described impressions as predominantly positive, which means that ChatGPT can support dialogic feedback and help to advance writing in different genres and micro parameters.

ChatGPT can also be used to create prompts to write, suggestions on how to revise, and how to eliminate grammatical mistakes as illustrated in multiple learning environments. Grammarly, Hemingway Editor and ChatGPT are redefining writing pedagogy (Grassini, 2023).

ChatGPT for Supporting Language Learners

ChatGPT is a multifunctional companion in language learning that can translation on the fly, simulate dialogues, and describe difficult grammar rules. It is a non-judgmental environment where second language learners find a safe outlet on which they can practice and experiment with language, hence building confidence. The fact that the model can create contextual examples and model real-life conversations perfectly fits the purpose of language learners, especially their oral and written communication (Ma et al., 2024).

Challenges in the Educational Use of ChatGPT

Despite the fact that ChatGPT can be of great benefit to language learning and grammar lessons, there are some issues as well. Such issues as excessive reliance on technology, inconsistent quality content, academic dishonesty, and inequitable opportunity are all included.

Effective inclusion in the learning process is dependent on the motivation of the learners, the availability of the technology, and its relevance to curriculum outcomes (Aljanabi & Alsalimy, 2024). One of the main arguments is possible over-reliance. Students can also be susceptible to relying on AI to perform modestly on grammatical tests, writing style recommendations, and other writing-related areas, whereas their thinking could be left unengaged, and hence, they are unable to learn and develop independently (Sallam et al., 2023).

Also, ChatGPT is a mostly language-generating model and not an education-oriented one. Therefore, it is rather emotionally and culturally implicit and neglects emotional intelligence and cultural sensitivity as important elements of effective instruction particularly in countries with diversities and multilingual populations such as Pakistan (Fuchs, 2023). Under-resourced settings also make the implementation challenging due to technological limitations. The digital divide is a material challenge in such a country as Pakistan, in which not all the schools in the state have any reliable connection to the internet, computer laboratories, or teachers who are knowledgeable about AI (Khan et al., 2024).

Lastly, the critics indicate that such AI tools as ChatGPT may sometimes give false or biased answers, so-called hallucinations, that can mislead learners (Bang & Taylor, 2023). Failure to evaluate digital content critically may result to the students accepting the erroneous content blindly, and thus, misconceptions are perpetuated.

Theoretical Framework

Second Language Acquisition Theories

Input Hypothesis

The Input Hypothesis presented by Krashen shows the significance of comprehensible input a little more than the level of proficiency a learner is working at, a foreign language has to be brought into the limelight in a way that goes beyond the level of proficiency that a learner is approaching. Talking about the option of learning with the ChatGPT enable learners to get access to the content that is unique to this theory. AI is also able to support progressive learning and effective learning process through offering grammar exercises and explanations relative to individual needs.

Interaction Hypothesis

In acquiring 2nd language the Interaction Hypothesis by Long emphasizes the need to have significant interaction. ChatGPT might facilitate this in which the learners are eager to define the meaning and to correct mistakes, as the feedback might be delivered in real-time and in a two-way dialogue. This interactive feature is likened to the application of the actual language and, therefore, makes the learning process both possible and interesting.

Connection to AI-Assisted Learning

As noted by Thomas et al. (2017), ChatGPT and other AI representations interactively contribute to the creation of immersive learning systems. These tools do not contradict the concepts of SLA either, as they support the ideas of interaction, situational learning, and immediate feedback (Williams and Clark, 2018). The AI with SLA is a new approach to language acquisition, or to be more exact, to learning English proficiency. The presence of personalization and instant responses of AI synthesis, as well as approach-based character of SLA in establishing its relations with the learner, simplify the process of learning and make it more pleasant. At this point, the research scope is developed based on these insights, which is the possible use of these AI-based tools, such as ChatGPT, along with conventional teaching to support language acquisition, thus, the basis of the research study as such.

Methodology

The study had a quasi-experimental research design in which two group were split consisting of 30 participants from each group i.e experimental and control group using convenient sampling technique. Both groups were taught by the same teacher (the researcher) and followed the same topic sequence over ten weeks. The experimental group received AI-supported instruction using ChatGPT, while the control group was taught through conventional method of teaching. The population for this study consisted of students of matriculation under the Federal Board of Intermediate and Secondary Education (FBISE) in Karachi, Pakistan. The target population was the Grade 10 students registered under Federal Board of Intermediate and Secondary Education (FBISE) in Karachi, Pakistan.

The research has been conducted in one of the private colleges in Karachi which was chosen as it was easily accessible and had the technology infrastructure. Two intact classes were taken for the study.

Data Collection Tools

The data were collected through pretest and posttest and a standardized scoring rubric was adopted. Moreover, t**he data was analysed through SPSS Version-26.**

Findings

Pre-Test of Control and Experimental Group for Overall Result

Both the control and experimental groups were assessed using the same 25-mark standardized grammar test.

Table 1

Descriptive Analysis of Overall Pre-Test Scores of Control and Experimental Group

Group	N	Mean	Std. Deviation	Std. Error Mean

Pre-Test – Overall Score	Control	30	10.80	1.424	.260
	Experimental	30	10.57	1.569	.286

Interpretation

As shown in the pre-test scores for the control group (M = 10.80, SD = 1.424) and experimental group (M = 10.57, SD = 1.569) in the descriptives, before the instructional changes were made, the control and experimental groups were close to the same level of English grammar knowledge. The group means vary by 0.23 points, and the group standard deviations were similar to indicate the same amount of change in scores from their means. These baseline scores allow for the two groups to be compared because of the equal opportunity to have baseline equal scores, which acts as a good starting point to measure the differences in learning outcomes from the control's traditional teaching methods to the experimental's teaching methods that were assisted by the applications of ChatGPT.

Table 2

Independent Samples Test of Overall Pre-Test Scores Comparing Control and Experimental Group

		Levene's Test for Equality of Variances		**t-test for Equality of Means**							
						Significance				**95% Confidence Interval of the Difference**	
		F	**Sig.**	**T**	**df**	**One-Sided p**	**Two-Sided p**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Pre-Test – Overall Score	Equal variances assumed	.146	.704	.603	58	.274	.549	.233	.387	-.541	1.008
	Equal variances not assumed			.603	57.464	.274	.549	.233	.387	-.541	1.008

Interpretation:

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

In regard to the first descriptive results, the control group average was 10.80 while the experimental average was 10.57 out of 25, showing almost the same score resulting in a 0.23 difference, which seemed negligible. An independent sample t-test was also conducted to compare both groups of students to determine if their scores in the grammar pre-test were equal to begin with.

In relation to the spread, both standard deviations of 1.42 and 1.57 were very low, indicating that the students in the groups were consistent in their performance. The two groups even had the same lowest score of 8 and the same highest score of 13, which suggests they were at the same starting point.

Post-Test of Control and Experimental Group for Overall Result

With Week 10 marking the end of the period of instructional interventions, the post-test was administered to assess the students' grammar proficiency. The two groups took the same 25-mark standardized test to ensure that the results were comparable.

Table 3

Descriptive Analysis of Overall Post-Test Scores of Control and Experimental Group

	Group	N	Mean	Std. Deviation	Std. Error Mean
Post-Test –Overall Scores	Control	30	16.27	1.461	0.267
	Experimental	30	18.63	1.245	0.227

Interpretation

There was a distinguishable show of improvement in the post-test results of the experimental group that utilised ChatGPT (M = 18.63) compared to the control group (M = 16.27) with a mean difference of 2.36. Additionally, the experimental group had a lower score variability (SD = 1.245) compared to the control group (SD = 1.461), indicating that the AI-assisted group had greater learning gains.

Table 4

Independent Samples Test of Overall Post-Test Scores Comparing Control and Experimental Group

	Levene's Test for Equality of Variances		**t-test for Equality of Means**							
									95% Confidence Interval of the Difference	
					Significance					
	F	**Sig.**	**t**	**df**	**One-Sided p**	**Two-Side**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**

							d p				
Post-Test Overall Scores	Equal variances assumes	1.776	0.188	-6.754	58	0.000	0.000	-2.367	0.350	-3.068	-1.665
	Equal variances not assumes			-6.754	56.584	0.000	0.000	-2.367	0.350	-3.068	-1.665

Interpretation:
Descriptive statistics indicate that the control group achieved an average score of 16.27 (SD = 1.46), with the lowest and highest scores of 14 and 19. This averages to an improvement of 5.47 points from the mean score of the pre-test (10.8), and it does suggest that traditional instruction has some positive impact on learning. On the other hand, the experimental group scored even higher with 18.63 (SD = 1.25), with the lowest score of 17 and the highest score of 21. This too indicates an 8.06 point increase from the pre-test (10.57), indicating that the experimental group made even more progress than the control group. To compare the post-test grammar proficiency scores of the control and experimental groups, an independent samples t-test was conducted. The results indicate that AI-assisted instruction via ChatGPT is more effective than traditional instructional methods in improving students' grammar proficiency.

Pre-Test and Post-Test Scores for Types of Sentences -Control Group

Table 5
Descriptive Statistics of Pre-Test and Post-Test Scores for Types of Sentences - Control Group

		Mean	N	Std. Deviation	Std. Error Mean
Control Group	Pre-test Score	3.4333	30	.62606	.11430
	Post-test Score	5.4333	30	1.04000	.18988

Interpretation
Those in the control group demonstrated a clear improvement in learning about the types of sentences after receiving traditional instruction. There was a gain of two points from the pre-test mean grade of 3.43 \([SD = 0.63]\) to the post-test mean grade of 5.43 \([SD = 1.04]\). The standard deviation was also higher in the post-test, showing a greater difference in the performance levels between the students.

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Table 6
Paired Samples Test of Pre-Test and Post-Test Scores for Types of Sentences -Control Group

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Control Group	Pre-test Score - Post-test Score	-2.00000	1.05045	.19179	-2.39225	-1.60775	-10.428	29	.000

Interpretation
There was a negative control group at first who only had traditional learning methods and this group also had a mean pre-test score at the sentence level of 3.43 (SD = 0.63). However, this mean post-test score jumped up to 5.43 (SD = 1.04). This shows a difference of 2 marks gained after the intervention, nullifying a null hypothesis as the control even improved. This suggests that, even though the group only had traditional learning methods, they still improved performance in identifying and constructing sentence types. With a paired samples t-test, results show the difference in pre-test and post-test scores to be statistically significant.

Pre-Test and Post-Test Scores for Types of Sentences – Experimental Group
Table 7
Descriptive Statistics of Pre-Test and Post-Test Scores for Types of Sentences - Experimental Group

		Mean	N	Std. Deviation	Std. Error Mean
Experimental Group	Pre-test score	3.6667	30	1.02833	.18775
	Post-test score	5.7000	30	1.20773	.22050

Interpretation
The group that utilised ChatGPT in class demonstrated improvements in their mastery of varying constructions such that their average score increased by a total of 2.03 points from 3.67 (SD = 1.03) in the test before the intervention to 5.70 (SD = 1.21) in the test after the intervention. Although the average score increase of the experimental group is slightly higher than that of the control group, the difference in starting means and in post-intervention means is evidence that the scores in ChatGPT-assisted students may have increased by a greater degree than the score increases demonstrated by students in the control group.

Table 8
Paired Samples Test of Pre-Test and Post-Test Scores for Types of Sentences - Experimental Group

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Experimental Group	Pre-test score Post-test score	-2.03333	1.40156	.25589	-2.55668	-1.50998	-7.946	29	.000

Interpretation

Participants in the experimental group achieved an average score of 3.67 (SD = 1.03) on the sentence-level grammar pre-test. This score improved to 5.70 (SD = 1.21) following the AI-assisted intervention. This improvement of 2.03 marks out of 8 is a bit higher than the improvement in the control group. Students in the experimental group appear to outperform students in the control group. The means of each score can be improved stepwise. The test of paired samples shows sufficient improvement in the post-test score compared to the pre-test score. t (29) = -7.95, p < .001 shows the means of each sample score improved to the 95% confidence limit (-2.56, -1.51). When considering the confidence limit, the improvement is proven to be of a significant order if not dramatic. The results are not due to random chance.

Pre-Test and Post-Test Scores for Phrases and Clauses – Control Group

Table 9

Descriptive Statistics of Pre-Test and Post-Test Scores for Phrases and Clauses – Control Group

		Mean	N	Std. Deviation	Std. Error Mean
Control Group	Pre-test Score	3.6000	30	.72397	.13218
	Post-test Score	5.3333	30	1.02833	.18775

Interpretation

Clauses and phrases showed an increase in the control group's performance due to classical subjects. They obtained an average score of 3.6 (SD=0.72) in the pre-test and 5.3 (SD=1.03) in the post-test, resulting in an average increase of 1.73. There was an increase in the standard deviation in post-test scores, indicating that there was a higher level of dispersion in student results following the teaching.

Table 10

Paired Samples Test of Pre-Test and Post-Test Scores for Phrases and Clauses – Control Group

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			

Control Group	Pre-test Score - Post-test Score	-1.73333	1.17248	.21406	-2.17115	-1.29552	-8.097	29	.000

Interpretation

The mean pre-test score for phrases and clauses in the control group stood at 3.60 (SD = 0.72), while the mean score at the post-test stage was 5.33 (SD = 1.03) after traditional teaching. The observed increase represents an average of 1.73 marks. The post-test scores, however, had a greater standard deviation, indicating that the students had a greater score spread, meaning that there were students who did not perform as well and who. The paired sample t-test confirmed that the observed improvement was not due to chance. It was evident that traditional instruction led to improvements and students in the control group were able to increase their skills and understanding of a phrase and clause.

Pre-Test and Post-Test Scores for Phrases and Clauses–Experimental Group

Table 11

Descriptive Statistics of Pre-Test and Post-Test Scores for Phrases and Clauses – Experimental Group

		Mean	N	Std. Deviation	Std. Error Mean
Experimental Group	Pre-test score	3.2000	30	.80516	.14700
	Post-test score	6.4000	30	1.03724	.18937

Interpretation

The improvement of the experimental group's understanding of phrases and clauses with ChatGPT assistance was pronounced. Enhancement was evidenced by mean score increases of 3.20 points from 3.20 (SD= .81) to 6.40 (SD=1.04). The improvement was considerably greater than that of the control group, implying the effectiveness of the AI tool towards the understanding of this concept of grammar.

Table 12

Paired Samples Test of Pre-Test and Post-Test Scores for Phrases and Clauses – Experimental Group

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Experimental Group	Pre-test score - Post-test score	-3.20000	.96132	.17551	-3.55896	-2.84104	-18.232	29	.000

Interpretation

The pre-test mean score of the experimental group was 3.20 (SD=0.81), indicating that the students had difficulty understanding the identification and use of phrases and clauses. Following 10 weeks of AI-assisted instruction with ChatGPT, the mean score of the post-test had increased to 6.40 (SD= 1.04), demonstrating that participants had improved by an average of 3.20 out of 8 and had mastered this aspect of grammar at a more advanced level. The paired samples t-test indicated that the pre-test and post-test samples measurably changed $t(29)= -18.23$, $p < 0.001$. The previous change was statistically significant. The ChatGPT characteristics of example generating, rule clarifying, and immediate feedback allowed students to more readily learn the complexities of phrases and clauses than traditional teaching. Overall, beyond the improved accuracy, the AI assistance and instruction had more evened out, increased, and more consistent performance of all learners, facilitating a decrease in the performance variability of the group.

Pre-Test and Post-Test Scores for Narrations – Control Group

Table 13

Descriptive Statistics of Pre-Test and Post-Test Scores for Narrations -Control Group

		Mean	N	Std. Deviation	Std. Error Mean
Control Group	Pre-test Score	3.7333	30	.73968	.13505
	Post-test Score	5.3333	30	1.06134	.19377

Interpretation

Through traditional instruction, the control group exhibited the following improvement in narrations. In the pre-test, they had a mean score of 3.73 (SD = 0.74), which increased to 5.33 (SD = 1.06) in the post-test. This is a mean improvement of 1.60 points. This standard deviation illustrates that the final scores of the students regarding the mastery of this topic were more heterogeneous.

Table 14

Paired Samples Test of Pre-Test and Post-Test Scores for Narrations - Control Group

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Control Group	Pre-test Score- Post-test Score	-1.60000	1.19193	.21762	-2.04507	-1.15493	-7.352	29	.000

Interpretation

Prior to the commencement of the traditional pedagogy instruction, the control cohort had a mean pre-test score of 3.73 (SD = 0.74), while following the traditional pedagogy instruction, the control cohort attained a post-test mean score of 5.33 (SD =

1.06). As a result of this, they increased by 1.60 points, which implies that traditional instruction in classroom activities, narration, and discussion exercises positively influenced students' narration skill. The t-test for dependent samples t(29) = -7.35, p < .001, with a 95% CI of -2.05 to -1.15 shows that narration improvement was statistically significant. The control group had teacher-centred pedagogy and somewhat increased the narration improvement in teacher-centred pedagogy. The improvement of 1.60 points was statistically significant but was less than what was expected in the improvement of the other grammar concepts such as sentence structure.

Pre-Test and Post-Test Scores for Narrations – Experimental Group

Table 15

Descriptive Statistics of Pre-Test and Post-Test Scores for Narrations –Experimental Group

		Mean	N	Std. Deviation	Std. Error Mean
Experimental Group	Pre-test score	3.6333	30	1.09807	.20048
	Post-test score	6.5333	30	1.38298	.25250

Interpretation

Participants in the experimental group showed pronounced improvement in the use of the tool ChatGPT in the construction of their narratives. The group means of their scores shifted from 3.63 (SD = 1.10) to 6.53 (SD = 1.38), representing a significant gain of 2.90 points. The results confirm that instruction supplemented with artificial intelligence is beneficial in supporting the participants' understanding of the construct and application of narrative grammar.

Table 16

Paired Samples Test of Pre-Test and Post-Test Scores for Narrations - Experimental Group

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Experimental Group	Pre-test score - Post-test score	-2.90000	1.72906	.31568	-3.54564	-2.25436	-9.186	29	.000

Interpretation

Specifically for the pre-test stage, the control group also posted a similar weak performance with a mean score of 3.84 (SD = 1.12), during the pre-test phase. This means the control group also weakly managed an average mean score of 3.00 points (out of 9) and so their post-test mean score also increased by the same amount, to 6.84 (SD = 1.19). This also indicated clear progression of the control group in managing intervening changes. The paired samples t-test, with an equal number of 30 students in each group, yielded a t-test score of 1.20, and a score of 9.89, with a two-tailed

probability of 9.16224 (-27.84 (1.91)). Equitable distribution also allows for near-universal improvement, post-AI direction. This means the group performance was indeed differentiated to near equalisation with the control group. AI-supported direction was also characterised by equity in the gap of outliers.

Discussion

At the outset, a placement test confirmed that the two groups began at comparable proficiency levels. Pre-test scores were nearly identical, although normality checks suggested departures from perfect normality, parametric analyses were still appropriate because of design robustness and consistent sampling.

Over the 10-week intervention, divergent patterns emerged. Both groups improved, but the magnitude of improvement was notably higher in the experimental group. Ongoing performance (Weekly quizzes) monitoring also showed that the experimental group consistently outperformed the control group in sentence construction, phrases and clauses, and narration**.**

The methodology played a decisive role here. Traditional instruction relied on teacher-led explanation, board work, and textbook exercises. In contrast, ChatGPT provided real-time, individualized support: students could test their own sentences, seek clarifications, and receive immediate corrective feedback. This interactive mode shifted learning from passive reception to active exploration, which aligns with the constructivist model of SLA (Ellis, 2020), and also resonates with Krashen's Input Hypothesis, since learners received clear input tailored to their immediate needs.

Overall effect of AI-assisted learning through ChatGPT on students' grammar proficiency compared to traditional teaching methods

Outcomes indicated improvement across all participants; nonetheless, the experimental group advanced even more so than the others. This was exemplified by the topic-specific quizzes and further solidified by the final post-test, indicating that the triad of abundant practice, immediate feedback, and learner autonomy allowed for the rapid learning of the material.

It is observed that in the AI-supported sessions, learners were more willing to "try and see." Students frequently drafted a sentence, checked it with ChatGPT, and then iteratively refined it. Engagement was generally high, partly because each learner could move at a personal pace while still being driven by shared objectives. Quieter students, who often remained silent during whole-class teaching, were notably more active when interacting with the tool.

Impact of ChatGPT on students' ability to identify and construct different sentence types

Sentence construction was assessed in Weeks 2–3 via quizzes and reinforced in the post-test. The experimental group showed higher improvement in post-test score. Students using ChatGPT could generate numerous examples of simple, compound, complex, and various functional types, then stress-test them by asking the tool to diagnose structure and explain the choice of conjunctions and punctuation. Repetition with variation, quickly producing multiple options, swapping clauses, or changing conjunctions, appeared to deepen pattern recognition beyond what time typically permits in board-based lessons. This process embodied Long's Interaction Hypothesis, as students negotiated meaning through dialogue with the tool, receiving

adjusted feedback in real time. In the beginning, students often confused compound with complex constructions or mishandled punctuation around coordinators. With ChatGPT, students experimented with different clause orderings and immediately checked explanation for why a certain sentence is simple, compound, or complex.

Effectiveness of ChatGPT in improving students' usage of phrases and clauses

Post-test revealed that the experimental group outperformed the control group in phrases and clauses section. The AI's ability to break down abstract grammar concepts into step-by-step explanations with contextual examples gave students more entry points to understanding. This reflects Krashen's Input Hypothesis, since repetitive, comprehensible input supported acquisition of abstract grammatical units

Recurrent practice with immediate explanations seemed particularly helpful for boundary detection (where the phrase or clause begins and ends), head–modifier relations, and the role of subordinators. Because the tool could instantly generate contrasting pairs (e.g., a noun phrase vs. a noun clause in parallel slots), students saw how form maps onto function across multiple contexts. Students often asked the model to turn a clause into a phrase and vice versa, and then looked for the exact change that accomplished the shift.

Impact of ChatGPT in improving students' understanding of narrations

The Weeks 7-8 assessment focused on Narration, a grammar aspect that instructors often find to be the most difficult. Also, the experimental group's post-test score in the narrations component was significantly higher than the control group. Here, ChatGPT was instrumental in providing practice at transformations where students changed sentences from direct to reported speech (and vice versa) with the help of the API. By pointing out students' mistakes, the AI was able to alleviate the fossilisation of their errors. When students tested ChatGPT with their sentences, they became very engaged and even tried it with different prompts like "What if the reporting verb changes?" or "What if the time marker is different?" Some students engaged in a mechanical 'rule hunting' process of conversion that stopped at explanation. In these cases, the teacher later moderated this by asking students to give a short explanation in every conversion (e.g. What tense changed and why?) or by asking for counterexamples to ensure they had understood. The quality of the transformations produced by the entire class improved over time in terms of clarity and there was also a greater consistency in the reasoning behind them.

Recommendations

Considering the findings from the research, important suggestions are made for the participants in the field of English language education. To begin with, educational technology such as ChatGPT ought to be incorporated as an additional grammar tool offering personalised exercises and feedback. Next, educators need to acquire additional training on how to use these tools. Then, there should be a transformation in the learning materials/service delivery from one that leans heavily on memorisation to one that is more active, using AI to enrich tasks based on the desired outcomes of the learning session. Then, there must be policy and infrastructural provisions to mitigate the digital gap that is characterised by the lack of adequate devices and proper IT with the digital education tools expensive necessary for education technologies. Then, to improve the learning of grammar, a more formal blend of

classroom instruction and AI-assisted learning is recommended.

Conclusion

This study shows that in the case of the instruction of grammar and composition for Pakistani ESL students, the use of ChatGPT resulted in statistically significant enhancements in grammatical competence, particularly in the domain of understanding various types of sentences, phrases and clauses, as well as in the writing of narratives. While the control group also benefited from instruction, the experimental group which received AI instruction gained significantly more with a post-test average of 18.63 and 16.27 for the control group. The study results also point to the potential of ChatGPT as a complement to more traditional instruction because of the ability to provide immediate feedback. The study suggests that educational institutions begin to use AI tools as part of a blended learning framework with sufficient pre-service and in-service teacher training, curriculum development, and equity of educational resources. The use of technology in education in a more thoughtful manner in the study also highlights the need to make the process of learning grammar in this kind of ESL situation more accessible, engaging and efficient.

References

Ali, A., Samoon, A. W., & Ali, M. (2020). STUDY OF DISTINGUISHING FEATURES OF PAKISTANI STANDARD ENGLISH. Elite English and Literature Journal, 7(2), 112. https://doi.org/10.24252/10.24252/elite.v7i2a1

Aljanabi, M. A. M., & Alsalimy, H. A. (2024). Exploring students' reliance on AI tools and its impact on learning autonomy. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12285-z

Baig, S., Javed, F., Siddiquah, A., & Khanam, A. (2021). A content analysis of English textbook of Punjab Textbook board of Grade 8 in Pakistan. SAGE Open, 11(2). https://doi.org/10.1177/21582440211023159

Bano, S. (2025). Pakistani Students Deserve Better: Why Grammar Translation Method Doesn't Work for English Learning. Education Library, 2(1), 50-61.

Bang, M., & Taylor, E. (2023). Hallucinations in large language models: A challenge for AI-powered education. Educational Technology Research and Development. https://doi.org/10.1007/s11423-023-10255-1

Fuchs, C. (2023). Ethical implications of using large language models in education. AI & Society. https://doi.org/10.1007/s00146-023-01656-5

Grassini, S. (2023). Shaping the Future of Education: Exploring the potential and consequences of AI and ChatGPT in educational settings. Education Sciences, 13(7), 692. https://doi.org/10.3390/educsci13070692

Haidar, S., & Fang, F. (2019). English language in education and globalization: a comparative analysis of the role of English in Pakistan and China. Asia Pacific Journal of Education, 39(2), 165–176. https://doi.org/10.1080/02188791.2019.1569892

Hussain, I., Ahmad, M., & Sajid, M. A. (2020). Digital competence and ICT integration in public sector secondary schools in Pakistan. Journal of Educational Research, 23(2), 34–50.

K, G., Goudar, R. H., M, D. G., Kaliwal, R. B., Rathod, V., Deshpande, S. L., Kulkarni, A., & Hukkeri, G. S. (2024). Enhancing Education with ChatGPT:

Revolutionizing Personalized Learning and Teacher Support. EAI Endorsed Transactions on Internet of Things, 10. https://doi.org/10.4108/eetiot.6998

Khan, M. S., & Ahmed, R. (2023). Technology integration in ESL classrooms: Challenges and opportunities in Pakistan. Journal of Education and Educational Development, 10(2), 45–60. https://doi.org/10.22555/joeed.v10i2.890

Khosa, M. S. (2021). Evaluation of sentence construction in English essays: A case study of Grade 10 learners of Malamulele Northeast Circuit (Doctoral dissertation).

Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. Computers and Education Artificial Intelligence, 7, 100278. https://doi.org/10.1016/j.caeai.2024.100278

Mohammad, N., Masum, R., Ali, Z., & Baksh, K. (2018). Teaching practices of English language in the schools of Lasbela District, Pakistan. International Journal of Experiential Learning & Case Studies, 2(2), 34–39. https://doi.org/10.22555/ijelcs.v2i2.1942.g349

Richards, J. C., & Rodgers, T. S. (2014). Approaches and Methods in Language Teaching (3rd ed.). Cambridge: Cambridge University Press.

Roberts, L., & Jones, D. (2019). Personalized Learning: A Comprehensive Approach. Journal of Educational Psychology, 111(4), 719-735.

Sajid, M., & Siddiqui, J. A. (2015). Lack of academic writing skills in English language at higher education level in Pakistan: causes, effects and remedies. International Journal of Language and Linguistics, 2(4), 174-186.

Tabassum, A. (2020). Teachers' self-efficacy and beliefs regarding grammar-translation method: Teaching writing at school level. http://dspace.bracu.ac.bd/xmlui/handle/10361/14352

Thomas, G., et al. (2017). Student Engagement and Blended Learning. Journal of Educational Research, 110(5), 446-459.

Shamim, F., & Tribble, C. (2005) — Language and education in Pakistan: Issues of access and equity.

Songsiengchai, S., Sereerat, B., & Watananimitgul, W. (2023). Leveraging Artificial Intelligence (AI): Chat GPT for Effective English Language Learning among Thai Students. English Language Teaching, 16(11), 68. https://doi.org/10.5539/elt.v16n11p68

Wasti, A. T. (2016). The Role of Literary Texts in Pakistani EFL Classrooms: Issues and challenges. http://repository.essex.ac.uk/17123/

Williams, J., & Clark, L. (2018). AI and Personalized Learning. Journal of Educational Technology, 47(3), 35-49

Yang, Y., Chen, L., He, W., Sun, D., & Salas-Pilco, S. Z. (2024). Artificial Intelligence for Enhancing Special Education for K-12: A Decade of Trends, Themes, and Global Insights (2013–2023). International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-024-00422-0

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895