Research Paper   Open Access

Empowering Grammatical Accuracy in Moroccan EFL Writing: The Impact of ChatGPT

Mourad Lahoual, Raha Yassine, Mohamed Ben Mhamed, YOUCEF *Hdouch*
[1](FLLA/ IbnTofail, Morocco)
[2](FLLA/ IbnTofail, Morocco, FLLA/ IbnTofail, Morocco/ FLLA/ IbnTofail, Morocco/ FLLA/ IbnTofail, Morocco)
Mourad Lahoual

ABSTRACT: This study examines the impact of AI-assisted tools, specifically ChatGPT (3.5 model), on enhancing grammatical accuracy and reducing errors in academic writing among Moroccan EFL university students. Employing a pretest-posttest control group design, 60 second-year economics students were randomly assigned to an experimental group (n=30) using ChatGPT paragraph revision, while the control group (n=30) wrote without AI assistance. Paired and independent samples T-tests revealed significant reductions in comma splices, run-on sentences, verb tense, and subject-verb agreement errors in the experimental group compared to the control group. On the other hand, the control group showed minimal, non-significant improvements. These findings highlight ChatGPT's efficacy in addressing linguistic challenges faced by Moroccan university students, mainly those who study English for Specific Purposes. This research underscored the vitality of integrating AI tools into EFL curricula and enhancing digital infrastructure in Moroccan education.

KEYWORDS - AI chatbots, ChatGPT, EFL writing, grammatical accuracy, writing mechanics, Moroccan university students

I.   INTRODUCTION

Academic writing is an important skill that English as a Foreign Language EFL university students dedicate significant effort to mastering, as it is the key to their success in academic life. Proficiency in academic writing is essential for EFL learners in their writing. Nonetheless, EFL learners frequently need support in achieving language accuracy and coherence. Following academic conventions in their writing so that they resort all type of machine assisted tools to ameliorate their academic writing proficiency Thus, since the release of ChatGPT (3,5 model) on November 30, 2022, Most of the EFL learners' resort to the artificial intelligence (AI) chatbot to develop the quality of their essays, research projects, and even their doctoral theses. Today, the AI chatbots are a magic tool that assists learners in finishing their assignments and projects quickly and with less effort. However, there is a controversial polemic about the ethicality of using AI chatbots. Academic integrity, along with cheating, is the serious issues that dominate the global academic discourse.

1.1. *Statement of the problem:*

Internationally, a great deal of research has demonstrated that AI chatbots significantly enhance EFL students' writing skills (Marzuki et al ,2023; Teng ,2023). However, Locally, the Moroccan context suffers from a dearth of empirical studies on the integration of AI chatbots in language teaching and learning. To address this gap, this study investigates the impact of AI chatbots on developing Moroccan EFL learners' writing skills, employing an experimental research design. To elaborate things more, most studies carried out in the Moroccan context have focused primarily on eliciting university EFL learners' perceptions and attitudes through qualitative or quantitative research designs or reviewing literature highlighting benefits and limitations of AI chatbots *(Bekou et al., 2024; Aherrahrou ,2024; Azennoud ,2024; Moussa and Belhaiah ,2024; Ouahani & Mahraj ,2025).* However, there remains a critical gap in experimental research that quantitively assesses the impact of AI-driven tools like ChatGPT on improving grammatical accuracy in academic writing among Moroccan university EFL learners.

1.2. *Significance of research:*

This study aims to fill that lacuna by providing empirical evidence on the measurable effects of ChatGPT on writing proficiency. Given the widespread adoption of AI tools among students, assessing their pedagogical efficacy based on scientific criteria has become compulsory, thereby enriching Moroccan

academia and enabling cross-contextual comparisons with global research trends. Moreover, these research results will supply valuable insights that can guide instructors and institutions in effectively integrating ChatGPT. This study can also offer valuable recommendations to all the stakeholders to cater to students' needs and to guide EFL learners towards the responsible, safe, and ethical use of all the AI-based writing tools.

1.3. *Research Objective:*

This research aims to accomplish this objective:

Research objective: to examine the impact of AI assisted tools in enhancing Moroccan university EFL learners' grammatical accuracy in Academic writing

1.4. *Research Question:*

This study addresses these research questions:

Research Question:

1. To what extent can AI assisted tools enhance Moroccan university EFL learners' writing accuracy in academic writing?
2. To what extent can AI assisted tools reduce students' grammatical mistakes and writing mechanic errors?

1.5. *Research Hypothesis:*

3. The research hypotheses originating from the above research question is

RH1: AI assisted tools positively enhance Moroccan university EFL learners' grammatical accuracy in academic writing.

RH2: AI assisted tools can reduce students' grammatical mistakes and writing mechanic errors.

II. REVIEW OF LITERATURE

2.1. AI Chatbots as Writing Assistants: Benefits and Challenges for EFL Learners

Writing is one of the most challenging skills students encounter in learning a second or foreign language. These difficulties ranged from vocabulary, grammar, and punctuation to expressing and organizing ideas, paragraph coherence, cohesion, and the fear of making mistakes (Sasmita & Setyowati, 2021). These constraints are triggered by inadequate resources, low English proficiency, and limited writing practice opportunities (Aldabbus & Almansouri, 2022).

To overcome these challenges, researchers suggested various strategies. For example, Ferris and Hedgcock (2014) recommended extra support for learners by engaging them in extra reading activities, brainstorming and generating ideas, using planning strategies, practicing drafting, and incorporating feedback, since ESL/EFL learners need more assistance than native learners. Sasmita and Setyowati (2021) proposed that teachers must provide additional sessions and provide their learners with positive feedback to enhance learners' self-esteem in writing skills (Ferris & Hedgcock, 2014; Hyland & Hyland, 2019)

Since launching the ChatGPT language generator AI machine in 2022, it has displayed significant potential in evolving EFL learners' writing skills, including brainstorming, outlining, improving vocabulary and grammar, and learning different writing styles. It provides interactive, real-time language practice and personalized and momentous feedback, enhancing student engagement and writing proficiency (Tlili et al., 2023). Nonetheless, opting for AI chatbots can raise learners' plagiarism, which may fade their creativity and critical thinking (Harunasari, 2023; Irzawati et al., 2024). Likewise, Carlson et al.(2023) pointed out that feedback generated by AI chatbots like ChatGPT may be complex, demotivating, and sometimes misleading.

AI chatbots have unquestionable potential and challenges, and their use has grown essential. Many researchers recommend the ethical use of AI tools as a result. Clear policies must be established by academics, legislators, and other stakeholders to guarantee the ethical application of AI. This will lessen its drawbacks, and EFL instructors and students will be encouraged to include these resources in their lesson plans. Teachers and students should improve their digital literacy or obtain ongoing, relevant training (Alsaedi, 2024).

2.2. The Effectiveness of AI Chatbots in Developing Writing Skills

Marzuki et al.'s (2023) research assesses the impact of AI tools on enhancing EFL students' writing quality from teachers' perspectives. Opting for a qualitative research design, the research revealed that integrating AI chatbots positively develops EFL students' writing quality in terms of content and organization.Teng (2023) used a mixed-method approach to investigate the impact of AI tools on Macu EFL students' writing. The results supported the significant positive effects of AI assistance on writing, incorporating writing motivation, self-efficacy, engagement, and a tendency to collaborate.

Allen and Mizumoto (2024) compared the effectiveness of writing groups versus AI technology (ChatGPT-3.5) for editing and proofreading academic writing among 33 Japanese EFL learners. This study revealed the students' preference for AI technology in editing and proofreading, though they acknowledged the efficiency of the writing groups. They emphasized the vitality of (ChatGPT-3.5) in giving effective, personalized, timely, and implementable feedback. Ghafouri et al. (2024) investigate the potential of positive

psychology interventions and artificial intelligence (AI) tools to enhance the psycho-emotional dimensions of second language (L2) teachers and learners. In a 10-week writing instruction program, the study concluded that in addition to the positive effect of ChatGPT-3.5 on teaching self-efficacy, EFL learners significantly improved their writing performance.

Most of the studies reviewed in this research found that EFL learners acknowledged the efficiency of the AI assistance feedback for their writing, emphasizing its efficiency as a tool for proofreading, editing, and enhancing their writing accuracy and cohesion. However, these studies highlighted the limitations of AI assistant writing tools regarding superficiality, creativity, and originality. Barrot (2023) pointed out that though AI tools can provide learners with efficient and timely feedback, they can never replace teachers' feedback. Thus, they should complete and supplement the machine's feedback. AI chatbots must be considered helpful friends.

2.3. Moroccan context

Amid the global rise of AI tools, these revolutionary chatbots have profoundly impacted academia in Morocco, particularly foreign language teaching. Many Moroccan researchers are investigating the effectiveness of AI tools in enhancing EFL learners' writing skills. Moussa and Belhaiah (2024) conducted a quasi-experimental study with Business Law undergraduates at the International University of Rabat, Morocco, to examine the relationship between EFL students and AI-assisted writing tools, focusing on writing skill development. The study revealed that AI tools significantly improved students' proficiency, creativity, organizational skills, vocabulary, grammar, and writing mechanics. Azennoud (2024) probed the efficiency of AI-assisted writing tools in enhancing university EFL students' writing accuracy and complexity. The research employed a quantitative design to achieve this objective by administering a Likert scale questionnaire addressed to university students. The study revealed that EFL learners. They started producing error-free and complex syntactic structures using AI-assisted writing chatbots.

Notably, most studies in the Moroccan context have focused primarily on eliciting university EFL learners' perceptions and attitudes through qualitative or quantitative research designs. However, there remains a significant gap in experimental research examining the impact of AI-driven tools on improving grammatical accuracy in academic writing among Moroccan university EFL learners.

III. METHODOLOGY

3.1. Research Design

This research adopted the Pretest-Posttest Control Group Design to explain the link between the causes and the effects of a defined research problem. Similarly, to better understand the link between the dependent variable, AI-assisted tools, and the independent variable, Moroccan university EFL grammatical writing in their academic writing, and to efficiently assess the impact of AI chatbots on Moroccan university EFL grammatical accuracy in their academic writing. Research design will likely guarantee validity, reliability, and replicability (Abbuhl et al.,2013). In addition to the wide use of experimental research in educational science, it is an efficient method to obtain the maximum amount of data from a small number of participants. Experimental design methods enable researchers to get insight and assess the factors that impact a particular system through a numerical statistical approach.

Researchers employed designed a well-defined rubric to ensure reliability, consistency and validity in scoring.

Category	Score 0 (Poor)	Score 1 (Fair)	Score 2 (Good)	Score 3 (Excellent)
Subject-Verb Agreement	Frequent errors (>5)	Some errors (3-5)	Few errors (1-2)	No errors
Sentence Fragments	Multiple incomplete sentences	1-2 fragments	Occasional fragments	No fragments
Run-on Sentences	Multiple run-ons (>3)	1-2 run-ons	Rare run-ons	No run-ons
Comma Splices	Frequent misuse (>4)	Some misuse (2-3)	Rare misuse (1)	No misuse
Verb Tense Errors	Consistent tense shifts	Occasional tense errors	Minor errors	No errors

▪ **Scoring Instructions**:
- Each category is scored **0–3**.
- Total score per paragraph: Sum of all categories (**Max = 18**).
- Two raters score blindly; calculate inter-rater reliability (**Cohen's d ≥ 0.80).**

3.2. Participants:

60 university students in the second-year major in economic studies in the faculty of the Legal, Economic and Social studies in Sale were recruited. Random assignment is employed in this research to ensure the efficient comparability of the groups.30 experimental participants used AI chatbots (Grammarly and Chatgpt 3.5version)

to correct their paragraphs. However, (n=30) were control group. They wrote their paragraphs without any AI assistance.

3.3. Design the Experiment:

Researchers employed Pretest-Posttest Control Group Design. The focused group and the control group were assigned to write a coherent paragraph. In the intervention phase, The focused group were asked to review their paragraphs with the assistance of AI associated tools (Chatgpt 3,5) and paying attention all the grammatical errors (subject verb agreement, fragment and run on errors, come splice…)

The researchers' intervention was done throughout the whole semester. In each session participants conducted a structured AI revision of their paragraph focusing on two types of errors. In the first session, learners concentrate only on subject-verb agreement and tense consistency. In the second session, they work only on prepositions and word orders. The Third session learners' focus was casted solely on the use of auxiliary verbs. In the fourth and fifth weeks, students prioritize run on and fragment sentences in their revision with assistance of AI tools. The last two weeks in the semester, the participants dealt only with punctuation errors including comma splice, missing periods. In the post-test, the two groups were asked to write another paragraph talking AI chatbots as an innovative technology that can assist students in their academic research.

3.4. Instruments:

3.4.1. *Chatgpt 3,5 model*: It is a free version was first launched by OpenAI as a free research preview on **November 30, 2022**, marking the public debut of its conversational AI capabilities. In this researcher, participants were exposed to a session in which they were provided with all the necessary guidelines to give suitable prompts.

3.4.2. *Writing tasks*: The participants were required to write paragraphs to assess the participants grammatical accuracy in their academic writing. They were asked to write 4 paragraphs on these topic sentences. This table presents the topics for the writing tasks.

Pretest paragraphs	Posttest Paragraphs
A person inspires you most.	A balanced diet is essential for overall well-being.
AI chatbots has advantages and drawbacks.	Social media platforms negatively impact the way people perceive their image.
Having a permanent contract in a multinational company has many limitations.	Reading is a fundamental skill that plays a crucial role in personal and intellectual development.
Studying abroad is very beneficial.	Hiking is a good exercise for adult.

3.5. Data Analysis:

To calculate the results for the study, the researchers implemented a paired samples T-test. The difference between the two groups was interpreted based on the mean scores, standard deviation (SD), and the level of significance of the difference between the groups according to the test results. In addition, the researcher compared the errors of the two groups to find which errors were reduced after using AI chatbots (model 3,5)

IV.   RESULTS

TABLE 1: Pretest and Posttest Descriptive Statistics for Experimental and Control Groups

Group	Test	Mean	N	Std. Deviation
Experimental	Pretest	4.07	30	2.49
Experimental	Posttest	1.63	30	1.65
Control	Pretest	4.47	30	2.67
Control	Posttest	4.10	30	2.37

TABLE 2: Paired Samples Test Results for Experimental and Control Groups

Group	Mean Difference	Std. Deviation	95% CI (Lower, Upper)	t	df	p-value	Cohen's d
Experimental	2.43	2.05	(1.67, 3.20)	6.52	29	<0.001	1.19
Control	0.37	2.39	(-0.52, 1.26)	0.84	29	0.407	0.15

Note. M Diff = mean difference (Pretest – Posttest), SD Diff = standard deviation of the difference, CI = confidence interval, df = degrees of freedom, d = Cohen's d (calculated as M Diff / SD Diff). Values are rounded to two decimal places, except p-values (three decimal places). $p < .001$ is reported as $p < .001$.

The experimental group, as analyzed in the paired t-test from" experimental group" showed a substantial and statistically significant improvement from pretest to posttest. The pretest mean was 4.0667, which decreased to 1.6333 on the posttest, producing a mean difference of 2.43333. This change was highly significant (t = 6.515, df = 29, p = 0.000), with a 95% confidence interval of [1.66944, 3.19723] that does not include zero, confirming the reliability of the difference. The effect size, calculated as Cohen's d, was approximately 1.19, indicating a large effect of the intervention.

In contrast, control group displayed minimal change from pretest to posttest, with no statistically significant improvement. The pretest mean was 4.4667, slightly decreasing to 4.1000 on the posttest, resulting in a mean difference of 0.36667. This difference was not significant (t = 0.842, df = 29, p = 0.407), and the 95% confidence interval [-0.52393, 1.25726] included zero, indicating no reliable change. The effect size, with a Cohen's d of approximately 0.15, was very small, which means the intervention had a weak impact.

Comparing the two groups, the experimental group exhibited a markedly greater development than the control group. The experimental group's mean difference (2.43333) was substantially larger than the control group's (0.36667), and its large effect size (d = 1.19) contrasted sharply with the control group's minimal effect (d = 0.15). The experimental group's highly significant p-value (0.000) versus the control group's non-significant p-value (0.407) further underscores the effectiveness of the intervention in the experimental group. These results suggest that the integration of the AI chatbots to the experimental group was highly effective in reducing the students fragment mistakes, while the control group experienced little to no meaningful change.

TABLE 3: Pretest and Posttest Descriptive Statistics for Run-on Sentence Errors in Experimental and Control Groups.

Group	Test	M	N	SD
Experimental	Pretest	1.73	30	1.17
	Posttest	0.80	30	0.85
Control	Pretest	2.67	30	2.02
	Posttest	1.83	30	1.86

TABLE 4: Paired T-Test Results for Pretest and Posttest Run-on Sentence Errors in Experimental and Control Groups.

V.

Group	*M* Diff	*SD* Diff	95% CI of Diff	*t*	*df*	*p*	*d*
Experimental	0.93	1.23	[0.47, 1.39]	4.16	29	<.001	0.76
Control	0.83	2.13	[0.04, 1.63]	2.14	29	.041	0.39

The results from **Table 3** and **Table 4**, which present the descriptive statistics and paired t-test outcomes for run-on sentence errors in the experimental and control groups, reveal distinct differences in the effectiveness of the adaptation of AI chatbot. According to **Table 3**, the experimental group's mean number of run-on sentence errors were dropped from 1.73 (*SD* = 1.17) on the pretest to 0.80 (*SD* = 0.85) on the posttest, reflecting a reduction of 0.93 points with a sample size of 30. The decrease in standard deviation from 1.17 to 0.85 displays a more consistent performance across participants post-intervention. In contrast, the control group's mean errors decreased from 2.67 (*SD* = 2.02) to 1.83 (*SD* = 1.86), a reduction of 0.84 points, also with a sample size of 30. The control group's standard deviation remained relatively high (from 2.02 to 1.86), indicating persistent variability in scores and less uniform improvement. The experimental group not only started with fewer errors (1.73 vs. 2.67) but also achieved a lower posttest mean (0.80 vs. 1.83), suggesting a more pronounced reduction in run-on sentence errors compared to the control group.

Table 4 further elucidates these results through paired t-test results, which show that the experimental group's mean difference of 0.93 (*SD* Diff = 1.23) was highly significant (*t*(29) = 4.16, *p* < .001), with a 95% confidence interval of [0.47, 1.39] that excludes zero, confirming the reliability of the reduction. The effect size, Cohen's d = 0.76, indicates a -large impact of the intervention, underscoring its substantial effect on reducing run-on sentence errors. Conversely, the control group's mean difference of 0.83 (*SD* Diff = 2.13) was only marginally significant (*t*(29) = 2.14, *p* = .041), with a 95% confidence interval of [0.04, 1.63] that includes values close to zero, suggesting less certainty in the consistency of the change. The control group's effect size, Cohen's d = 0.39, reflects a small-to-moderate effect, indicating a weaker impact compared to the experimental group. The higher standard deviation of the difference (2.13 vs. 1.23) in the control group further points to greater variability in the change scores, reinforcing the experimental group's more robust and consistent improvement.

The experimental group demonstrated a greater and more reliable reduction in run-on sentence errors than the control group, as evidenced by a slightly larger mean difference (0.93 vs. 0.83), a highly significant p-value (< .001 vs. .041), and a nearly twice as large effect size (*d* = 0.76 vs. *d* = 0.39). The experimental group's tighter confidence interval and lower variability in change scores suggest that the use of ChatGPT was effective and consistently applied across participants. In contrast, the control group's marginal significance and wider confidence interval indicate that any improvement may be less dependable and potentially attributable to factors other than a structured intervention, such as practice effects or random variation. These findings suggest that the AI implemented in the experimental group was markedly more effective in reducing run-on sentence errors, producing a substantial and statistically significant improvement in comparison to the weak change observed in the control group.

TABLE 5: **Pretest and Posttest Descriptive Statistics for Comma Splice Errors in Experimental and Control Groups.**

Group	*M* Diff	*SD* Diff	95% CI of Diff	*t*	*df*	*p*	*d*
Experimental	2.43	2.05	[1.67, 3.20]	6.52	29	<.001	1.19
Control	0.37	1.40	[-0.16, 0.89]	1.43	29	.163	0.26

TABLE 6: Paired t-Test Results for Pretest and Posttest Comma Splice Errors in Experimental and Control Groups

Table 5 presents the pretest and posttest means and standard deviations for comma splice errors. The experimental group's mean decreased from 4.07 to 1.63, while the control group's mean decreased slightly from 4.63 to 4.27. As shown in Table 2, the experimental group exhibited a significant reduction in errors (M Diff = 2.43, t (29) = 6.52, p < .001, d = 1.19), with a large effect size. In contrast, the control group's reduction was not significant (M Diff = 0.37, t(29) = 1.43, p = .163, d = 0.26), indicating a small effect. These results suggest that the intervention was highly effective in reducing comma splice errors in the experimental group, while the control group showed no meaningful improvement.

Table 6 explains the paired t-test results for comma splice errors, highlighting the intervention's impact. The experimental group showed a significant reduction in errors (M Diff = 2.43, t(29) = 6.52, p < .001, d = 1.19), with a large effect size, indicating a strong effect. However, the control group's reduction was non-significant (M Diff = 0.37, t(29) = 1.43, p = .163, d = 0.26), with a small effect, which suggests minimal change. The experimental

Group	Test	*M*	*N*	*SD*
Experimental	Pretest	4.07	30	2.49
	Posttest	1.63	30	1.65
Control	Pretest	4.63	30	2.22
	Posttest	4.27	30	2.30

group's tighter confidence interval [1.67, 3.20] versus the control's [-0.16, 0.89] underscores the intervention's reliable effectiveness.

The experimental group meaningfully outperformed the control group in reducing comma splice errors, with a mean difference of 2.43 (t (29) = 6.52, p < .001, d = 1.19) compared to the control's 0.37 (t (29) = 1.43, p = .163, d = 0.26). The experimental group's large effect size and tight 95% CI [1.67, 3.20] versus the control's non-significant result and wider CI [-0.16, 0.89] highlight the intervention's robust impact.

TABLE 7: Pretest and Posttest Descriptive Statistics for Verb Tense Errors in Experimental and Control Groups

Group	Test	*M*	*N*	*SD*
Experimental	Pretest	2.20	30	1.47
	Posttest	1.33	30	1.47
Control	Pretest	3.07	30	2.23
	Posttest	2.80	30	2.14

TABLE 8: **Paired t-Test Results for Pretest and Posttest Verb Tense Errors in Experimental and Control Groups.**

Group	*M* Diff	*SD* Diff	95% CI of Diff	*t*	*df*	*p*	*d*
Experimental	0.87	1.31	[0.38, 1.35]	3.64	29	.001	0.66
Control	0.27	1.08	[-0.14, 0.67]	1.35	29	.187	0.25

Table 7 and Table 8 present the descriptive statistics and paired t-test results for verb tense errors in the experimental and control groups. Table 7 shows that the experimental group's mean errors decreased from 2.20 (SD = 1.47) to 1.33 (SD = 1.47, N = 30), while the control group's mean errors dropped slightly from 3.07 (SD = 2.23) to 2.80 (SD = 2.14, N = 30).

Table 8 reveals that the experimental group's reduction (M Diff = 0.87) was significant (t (29) = 3.64, p = .001, d = 0.66), with a moderate-to-large effect and a 95% CI [0.38, 1.35] excluding zero. In contrast, the control group's reduction (M Diff = 0.27) was non-significant (t (29) = 1.35, p = .187, d = 0.25), with a small effect and a 95% CI [-0.14, 0.67] including zero. Comparing the groups, the experimental group achieved a larger, significant reduction in verb tense errors (0.87 vs. 0.27) with a greater effect size (0.66 vs. 0.25), indicating the intervention's effectiveness, while the control group's small change suggests no meaningful improvement.

TABLE 9: Pretest and Posttest Descriptive Statistics for Subject-Verb Agreement Errors in Experimental and Control Groups

Group	Test	*M*	*N*	*SD*
Experimental	Pretest	2.13	30	1.50
	Posttest	1.17	30	1.21
Control	Pretest	3.73	30	2.72
	Posttest	3.00	30	2.55

TABLE 10: Paired T-Test Results for Pretest and Posttest Subject-Verb Agreement Errors in Experimental and Control Groups

Group	*M* Diff	*SD* Diff	95% CI of Diff	*T*	*df*	*p*	*d*
Experimental	0.97	1.10	[0.56, 1.38]	4.82	29	<.001	0.88
Control	0.73	2.07	[-0.04, 1.51]	1.94	29	.062	0.35

Table 9 shows that the experimental group's mean subject-verb agreement errors decreased from 2.13 (*SD* = 1.50) to 1.17 (*SD* = 1.21, *N* = 30), while the control group's mean errors dropped from 3.73 (*SD* = 2.72) to 3.00 (*SD* = 2.55, *N* = 30). Table 10 indicates that the experimental group's reduction (*M* Diff = 0.97) was highly significant ($t(29) = 4.82$, $p < .001$, $d = 0.97 / 1.10 \approx 0.88$), with a large effect size and a 95% CI [0.56, 1.38] excluding zero, confirming a robust intervention effect. The control group's reduction (*M* Diff = 0.73) was not significant ($t(29) = 1.94$, $p = .062$, $d = 0.73 / 2.07 \approx 0.35$), with a small-to-moderate effect and a 95% CI [-0.04, 1.51] including zero, suggesting no reliable improvement.

Comparing the groups, the experimental group achieved a larger reduction (0.97 vs. 0.73), a highly significant result ($p < .001$ vs. $p = .062$), and a much larger effect size ($d = 0.88$ vs. $d = 0.35$), demonstrating the intervention's effectiveness. The experimental group's lower posttest mean (1.17 vs. 3.00) and tighter CI underscore a consistent impact, while the control group's non-significant change and wider CI suggest minimal progress, likely due to the absence of intervention. The AI integration in correcting students subject-verb agreement, thus, significantly reduced subject-verb agreement errors in the experimental group, whereas the control group also showed

Group	*N*	*M*	*SD*	*t*	*df*	*p*	Mean Difference	95% CI [Lower, Upper]	Cohen's *d*
Control (No ChatGPT)	30	8.14	2.22	-12.32	57	< .001	-7.03	[-8.17, -5.89]	3.21
Experimental (ChatGPT)	30	15.17	2.17						

meaningful improvement.

TABLE 11: Independent Samples T-Test Comparing Academic Writing Accuracy Between ChatGPT-Assisted and Control Groups

Note: Accuracy scores are based on a 20-point rubric assessing students' paragraphs. Equal variances were assumed (Levene's F = 0.094, p = .760). Cohen's d was calculated using the pooled standard deviation.

Researchers adopted independent samples T-test to compare the academic writing accuracy of the students using AI assisted tools (experimental group, n = 30) versus those using traditional methods (control group, n = 30). The AI assisted tools assisted the experimental group (M = 15.17, SD = 2.17) significantly outdid the control group (M = 8.14, SD = 2.22), t (57) = -12.32, p < 0.001. The mean difference was 7.03 points (95% CI: -8.17, -5.89), which represents 35% improvement on a 20-point scale. The effect size was large (Cohen's d = 3.21), indicating a substantial impact of AI chatbots on students' writing accuracy.

V. DISCUSSION

This research found that AI chatbots can assist Moroccan university students in significantly reducing their grammatical and writing mechanics errors and enhancing their academic writing accuracy. The independent samples T-test (Table 11) displayed that the experimental group obtained a significantly higher mean accuracy score (M =

15.17, SD = 2.17) comparatively with the control group (M = 8.14, SD = 2.22), t (57) = -12.32, p < .001, with a large effect size (Cohen's d = 3.21). Additionally, Paired T-tests provided a more detailed analysis of this effect, revealing significant reduction in particular error types—such as comma splices, run-on sentences, verb tense errors, and subject-verb agreement errors—in the experimental group, with effect sizes ranging from moderate to large (d = 0.66–1.19). Conversely, the control group showed minimal or non-significant improvements, underscoring the effectiveness of ChatGPT in enhancing students' writing accuracy and proficiency.

This study concluded that AI chatbots, namely ChatGPT, play an effective role in reducing specific grammatical and writing mechanical errors. For comma splice errors, the experimental group's mean error count

dropped significantly from 4.07 to 1.63 (t(29) = 6.52, p < .001, d = 1.19), while the control group's decrease was non- significant (t(29) = 1.43, p = .163, d = 0.26). Likewise, run-on sentence errors decreased from 1.73 to 0.80 in the experimental group (t (29) = 4.16, p < .001, d = 0.76), compared to a marginally significant reduction in the control group (t(29) = 2.14, p = .041, d = 0.39). Verb tense errors and subject-verb agreement errors also showed significant reductions in the experimental group (d = 0.66 and 0.88, respectively), with non-significant changes in the control group (p = .187 and .062). These findings align with many other studies. For example, Dodigovic (2007) found that AI is an efficient instrument of error remediation, reducing the error rate by an average of 83%. Hussain et al.'s (2024) study revealed that ChatGPT and Meta AI can effectively detect morphosyntactic errors. Meta AI offers a broader range of corrections encompassing grammar and style, enhancing readability and contextual clarity.

It is also found that the large effect sizes for comma splices (d = 1.19) and subject-verb agreement (d = 0.88) are particularly noteworthy since such errors are prevalent among Moroccan learners. The prevalence of such errors is attributed to the linguistic interference from Arabic, which lacks similar syntactic structures (e.g., overt subject-verb agreement markers). Therefore, AI instruments such as ChatGPT will likely address this challenge for Moroccan learners, as this tool can supply real-time, context-sensitive corrections.

The enormous effect size (d = 3.21) in the independent samples T-test indicates that ChatGPT appears particularly beneficial for Moroccan (EFL) learners, perhaps due to their dependence on external tools for addressing competency gaps. This corresponds to studies on EFL education in the Middle East, where AI technologies have been successfully utilized to correct Arabic-related errors (Alharbi, 2023). The robustness of the gains across error types, as shown by the narrow confidence intervals (e.g.,[1.67,3.20] for comma splices), provides additional evidence of the reliability of ChatGPT's effect, and therefore the potential of ChatGPT as a tool across Moroccan classrooms.

There are important implications for embedding AI-supported tools into Moroccan EFL education. First, ChatGPT could be integrated into writing curricula as a revision aid, allowing students to keep error rates low in real time without cutting back on valuable higher-order skills such as argumentation and coherence. Teachers can teach students to make much of the use of ChatGPT, clarify information, accept the correction, and expand their knowledge on the chat, for long-term learning (Fitria, 2021). Second, the strong effect of size across error types supports the case for investing in digital infrastructure, especially in rural Moroccan schools, to ensure access to AI tools for everyone. Teacher training programs should also focus on AI literacy, preparing the teachers to use tools like ChatGPT to meet educational objectives and deal with ethical challenges, such as dependence or academic probity.

At a policy level, the Moroccan Ministry of Higher Education and scientific and technical research could consider piloting the inclusion of AI tools in the EFL curriculum and harnessing the power of scale to address EFL teacher shortages and large classroom sizes. As identified in this study, a marked reduction in errors could mean that such programmes would raise national proficiency in English, mainly in classes that taught English as English for Specific Purpose (ESP).

Despite the validity and reliability of these research findings, many limitations must be highlighted. This study recruited only (N = 60) university students, which did not allow the researchers to generalize the research findings. Future research should enlarge the sample size and include students from other departments to ensure the generalizability of the research findings. Furthermore, this study concentrated only on specific error types (comma splices, run-on sentences, verb tense, subject-verb agreement). Nonetheless, other aspects of writing accuracy, such as lexical choice or coherence, were not taken into consideration. Thus, analyzing such characteristics in academic writing will give researchers and other stakeholders a comprehensive perspective on the efficiency of AI-based tools in developing Moroccan EFL students' accuracy in academic writing. Researchers resorted to paragraph writing as a test to assess ChatGPT's efficiency in reducing university students' grammatical and writing mechanical errors and enhancing their accuracy and proficiency. However, complex tasks like essay writing enable researchers to generate robust results as learners can yield as many error patterns as they must be examined.

VI.   CONCLUSION

This research makes a strong case that ChatGPT successfully reduces grammatical and writing mechanics errors in Moroccan EFL students, with large effect sizes, and drops consistently occurring in comma splices, run-on sentences, verb tense, and subject-verb agreement. The results indicate that AI offers possibilities to revolutionize EFL writing instruction in Morocco by solving linguistic and resource-related issues. Effectively woven into pedagogical and policy initiatives, ChatGPT and many other chatbots can enrich university students' experiences and support long-term learning. Further research and prudent application will help to ensure that the new AI tools deliver the greatest good and limit possible harm, leading, in turn, to more effective EFL teaching in higher education, either in Morocco or elsewhere.

VII. ACKNOWLEDGEMENTS

REFERENCES

[1]. Marzuki, Widiati U, Rusdin D, Darwin, Indrawati I. The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. Cogent Education. 2023 Dec 11;10(2):2236469.

[2] Teng MF. Scientific writing, reviewing, and editing for open-access TESOL journals: The role of ChatGPT. International Journal of TESOL Studies. 2023;5(1):87-91.

[3] Bekou A, Mhamed MB, Assissou K. Exploring opportunities and challenges of using ChatGPT in English language teaching (ELT) in Morocco. Focus on ELT Journal. 2024 Mar 5;6(1):87-106.

[4] Aherrahrou M. Assessing the English Language Lecturers' Skill in Identifying Originality in Essay Writing Among Moroccan S2 Students. InConference Proceedings. The Future of Education 2024 2024 Jun 21.

[5] Azennoud A. Enhancing Writing Accuracy and Complexity through AI-Assisted Tools among Moroccan EFL University Learners. International Journal of Linguistics and Translation Studies. 2024 Nov 17;5(4):211-26.

[6] Moussa A, Belhiah H. Beyond Syntax: Exploring Moroccan Undergraduate EFL Learners' Engagement with AI-Assisted Writing. Arab World English Journal (AWEJ) Special Issue on ChatGPT. 2024 Apr 2.

[7] Ouahani NE, MAHRAJ M. ChatGPT Application in ELT: the International Context, the Moroccan Context, and Future Prospects.

[8] Sasmita YV, Setyowati L. Problems faced by EFL students in learning to write. Linguista: Jurnal Ilmiah Bahasa, Sastra, Dan Pembelajarannya. 2021 Jun 21;5(1):11-25.

[9] Aldabbus S, Almansouri E. Academic writing difficulties encountered by university EFL learners. British journal of English linguistics. 2022;10(3):1-1.

[10] Ferris DR, Hedgcock JS. Teaching L2 composition: Purpose, process, and practice. Routledge; 2023 Mar 23.

[11] Hyland K, Hyland F, editors. Feedback in second language writing: Contexts and issues. Cambridge university press; 2019 Jul 4.

[12] Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, Agyemang B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart learning environments. 2023 Feb 22;10(1):15.

[13] Harunasari SY. Examining the effectiveness of AI-integrated approach in EFL writing: A case of ChatGPT. International Journal of Progressive Sciences and Technology (IJPSAT). 2023;39(2):357-68.

[14] Irzawati I, Unamo AF, Natasha K. The role of Twitter in Gen Z English Development. Esteem Journal of English Education Study Programme. 2024 Jun 9;7(2):485-94.

[15] Carlson JA, Cheng RZ, Lange A, Nagalakshmi N, Rabets J, Shah T, Sindhwani P. Accuracy and readability of artificial intelligence chatbot responses to vasectomy-related questions: public beware. Cureus. 2024 Aug 28;16(8).

[16] Alsaedi N. ChatGPT and EFL/ESL writing: A systematic review of advantages and challenges. English Language Teaching. 2024;17(5):41-50.

[17] Allen TJ, Mizumoto A. ChatGPT over my friends: Japanese English-as-a-Foreign-Language learners' preferences for editing and proofreading strategies. RELC Journal. 2024 Jul:00336882241262533.

[18] Ghafouri M. ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. System. 2024 Feb 1;120:103209.

[19] Barrot JS. Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing. 2023 Jul 1;57:100745.

[20] Abbuhl R, Gass S, Mackey A. Experimental research design. Research methods in linguistics. 2013;1:116-34.

[21] Dodigovic M. Artificial intelligence and second language learning: An efficient approach to error remediation. Language Awareness. 2007 May 15;16(2):99-113.

[22] Hussain F, Latif F, Mujtaba M, Sohail M, Latif F. Comparing the Effectiveness of Artificial Intelligence (AI) Models, ChatGPT and Meta AI to Detect and Correct Morphosyntactic Errors in BS Accounting Student's Written Text. Contemporary Journal of Social Science Review. 2024 Nov 19;2(04):700-14.