Original Article

Enhancing the Writing Proficiency of Non-English Major Students Using Artificial Intelligence (AI) Language Tools

Michael Garcia [1,*]

Received: 27 April 2026; Revised: 16 May 2026;
Accepted: 04 June 2026; Published: 09 June 2026

DOI: https://doi.org/10.66074/NET55E77R6

Abstract

The rapid adoption of Artificial Intelligence language tools has fundamentally transformed educational practices, creating a critical need to evaluate their specific impact on the higher-order writing skills of non-English-major college students. This study examined the writing proficiency reflected in the written outputs of non-English major students before and after AI-assisted revision across five distinct dimensions and tested for significant differences to propose a targeted instructional intervention program. Utilizing a quasi-experimental single-group pretest-posttest design, researchers evaluated a purposively selected sample of seventy-seven first-year students enrolled in a general education course. Participants composed an initial unassisted academic essay and a subsequently revised essay guided by Artificial Intelligence feedback, which were then scored using an expert-validated analytic rubric and analyzed using repeated measures ANOVA. The AI-assisted revision intervention was associated with statistically significant improvements in the quality of respondents' written outputs, elevating the overall mean score from a baseline of 25.55 to 29.97 with a substantial effect size. Despite these gains, originality and creativity, critical thinking, and tone and style remained at the emerging proficiency level, along with contextual understanding, which achieved a proficient mean but still saw the majority of the population at emerging proficiency. However, the observed improvements reflect short-term performance within AI-assisted revision tasks and should not be interpreted as conclusive evidence of permanent

[1] Pangasinan State University, School of Advanced Studies, Philippines
* michaelgarcia @ucu.edu.ph

Volume 2, Issue 2, June 2026

independent writing mastery. These findings suggest that while Artificial Intelligence operates as a highly effective structural scaffold, it cannot automatically cultivate the deep analytical and creative capabilities required for authentic scholarly authorship. To address these observed instructional gaps, the study proposes the theoretically grounded Project A.I.-V.O.I.C.E. framework as a potential intervention model designed to strengthen evaluative judgment and maintain human agency during AI-assisted writing tasks. However, the framework itself requires separate empirical validation in future research contexts.

Keywords: AI language tools, augmented intelligence, non-English major students, quasi-experimental study, writing proficiency assessment

1. Introduction

The rapid advancement of Artificial Intelligence (AI) has fundamentally transformed educational practices, particularly in how students approach the craft of writing. AI writing tools—ranging from generative platforms like ChatGPT and Gemini to specialized assistants like Grammarly and QuillBot—now serve as essential digital supports for generating ideas and improving grammar, clarity, and coherence. In this modern context, writing proficiency is defined as the ability to communicate ideas effectively through well-organized, mechanically accurate, and contextually appropriate writing. Within AI-assisted educational environments, students increasingly engage with digital tools that support idea generation, revision, and linguistic refinement during the writing process. Within the context of this study, writing proficiency refers specifically to the quality of respondents' written outputs as evaluated through the analytic proficiency rubric during AI-assisted revision tasks and should not be interpreted as conclusive evidence of permanent independent writing mastery. As these technologies become deeply embedded in the academic workflow, the educational community is shifting toward a model of "augmented intelligence," where AI is used to enhance rather than replace human cognitive performance.

Global and local data underscore the magnitude of this shift. Recent reports indicate that 86% of students utilize AI for academic tasks, with generative AI becoming a staple for assessments despite a lack of formal training (Ahmed, 2024; Freeman, 2025). In the Philippines, this adoption is mirrored across higher education, where students frequently leverage tools to bridge linguistic gaps, though they express persistent concerns regarding overreliance and originality (Garcia et al., 2025; Espartinez, 2024). While empirical evidence suggests AI can improve organizational structure and grammatical accuracy, it also highlights a critical "agency gap." Students who uncritically accept AI output often produce work that lacks the depth and

interrogation found in human-led revision, leading to ethical uncertainties and inconsistent institutional policies (Mekheimer, 2025; Archana et al., 2025).

Despite the growing body of research on AI adoption, a significant gap persists in understanding the qualitative impact of these tools on specific writing dimensions. Most studies focus on broad mechanical improvements, leaving a void in our understanding of how AI influences tone, style, contextual understanding, emotional resonance, and critical thinking. Furthermore, there is a lack of targeted research on non-English majors—individuals whose curricula may not prioritize advanced writing instruction but who require high-level writing skills for professional success. This lack of a nuanced problem statement leaves educators without a clear roadmap for supporting students who are linguistically vulnerable in an AI-saturated academic environment.

To address these gaps, this study investigated the role of AI language tools in supporting the writing proficiency reflected in the written outputs of non-English-major students during AI-assisted revision tasks. Specifically, the inquiry focused on determining the level of writing proficiency among student-respondents before and after AI assistance across five key dimensions: (a) tone and style, (b) contextual understanding, (c) emotional resonance, (d) critical thinking, and (e) originality and creativity. By examining whether significant differences emerge in students' writing performance following AI-assisted revision, the research aimed to pinpoint which areas of AI-enhanced writing require the most urgent instructional support. Central to this investigation was testing the null hypothesis at the 0.05 level of significance that there is a no significant enhancement in the writing proficiency of non-English major students before and after using AI language tools across the five identified dimensions.

The rationale behind this research lies in its potential to redefine AI as an instructional bridge rather than a substitute for learning. By identifying specific writing gaps, the study provides the empirical foundation for a proposed intervention program—Project A.I.-V.O.I.C.E.—designed to be both remedial and future-oriented. This approach matters because it equips students with essential AI literacy while correcting fundamental writing issues, ensuring that technological innovation does not come at the expense of academic integrity or human agency. Ultimately, this research connects the immediate needs of non-English major students to the broader goal of producing graduates who can navigate the digital world with both technical proficiency and rhetorical depth.

2. Methodology

2.1 Research Design

This study employed a quantitative, quasi-experimental single-group pretest-posttest design to examine changes in writing proficiency following AI-assisted revision among non-English major students. A quasi-experimental design is used to examine the effect of an intervention on a dependent variable without full random

assignment, which is particularly suitable for educational settings that use intact classes (Creswell & Creswell, 2018; Fraenkel et al., 2011). By measuring the same respondents before and after the intervention, the design allows for a direct comparison of individual improvement. It produces inferential quantitative data to identify significant shifts in performance. Because the study focused specifically on respondents with lower initial writing proficiency, the researcher recognizes the possibility of regression-to-the-mean effects, wherein unusually low pretest scores may naturally shift upward during subsequent testing.

However, the study was designed primarily as an exploratory educational intervention to examine patterns of improvement following AI-assisted revision among academically vulnerable learners, rather than to establish generalizable causal conclusions across all student populations. Although single-group pretest-posttest designs possess weaker internal validity than randomized controlled experiments, they remain widely accepted in educational research, where intact classes and authentic classroom interventions limit random assignment (Creswell & Creswell, 2018; Fraenkel et al., 2011; Shadish et al., 2002). Consequently, the study does not claim absolute causality but instead investigates whether statistically significant improvements occurred following AI-assisted revision. The researcher also recognizes that factors such as repeated task exposure, revision familiarity, and practice effects may have partially contributed to the observed improvements.

2.2 Research Locale and Participants

The study was conducted at Urdaneta City University (UCU), Pangasinan, during the second semester of Academic Year 2025–2026. The population consisted of 169 first-year non-English major students from five programs: Tourism Management (BSTM), Hospitality Management (BSHM), Accountancy (BSA), Physical Education (BPED), and Culture and Arts Education (BCAED). This study utilized purposive, criterion-based sampling. Following the pretest, only respondents at the “Beginning” to “Emerging” proficiency levels were selected for the intervention and post-test phases. Consequently, 77 respondents proceeded to the post-test and subsequent analysis. This sampling procedure was intentionally employed because the study specifically aimed to examine patterns of improvement among non-English-major students with lower writing proficiency in an AI-assisted revision context.

2.3 Instruments

The primary instruments were researcher-developed tools comprising a writing task and an analytic rubric. These instruments were validated by six experts in writing and technology, achieving a composite weighted mean of 4.63, interpreted as "Very Highly Valid." The task required a 250–500-word academic essay on gender-sensitive communication, aligning with the students' midterm curriculum. To assess proficiency, a five-component analytic rubric was used to measure: (a) tone and style,

(b) contextual understanding, (c) emotional resonance, (d) critical thinking, and (e) originality and creativity.

To further support scoring consistency, adviser moderation was conducted to review rubric application and evaluate consistency across the assessed essays. While standardized scoring criteria were used throughout the evaluation process, future studies may further strengthen methodological rigor by including blind scoring procedures, multiple independent raters, and inter-rater reliability testing.

2.4 Data Collection Procedure

The data-gathering procedure followed a systematic, chronological sequence, beginning with the acquisition of necessary administrative and ethical permissions. Prior to the study's commencement, the researcher secured a formal Ethics Clearance and transmitted request letters to the respective authorities, including the Executive Director of the Pangasinan State University School of Advanced Studies and the Acting President of Urdaneta City University.

Once approvals were granted, a pre-test was administered in which 169 students composed a 250–500-word essay within 1 hour without any AI assistance. Following the pre-test, all essays were evaluated using the validated five-component analytic rubric to establish baseline proficiency. Only those students identified as performing at the Beginning to Emerging levels were selected to proceed to the post-test phase.

During the intervention phase, selected respondents were given 1 hour to interact with AI language platforms, such as ChatGPT and Google Gemini, while revising their essays. Rather than merely editing their drafts independently, students actively engaged with AI-generated rubric-aligned feedback, structural recommendations, contextual refinements, grammatical corrections, and rhetorical suggestions. The revision process therefore, involved guided AI-assisted scaffolding rather than conventional unaided revision alone.

Within this process, respondents critically examined the AI-generated suggestions and selectively incorporated revisions they deemed appropriate. This integrated interaction ensured that the final outputs reflected not only technological assistance but also the students' evaluative judgment, independent decision-making, and active cognitive engagement throughout the revision process.

2.5 Data Analysis

The collected data were treated using both descriptive and inferential statistical analyses to address the research questions with precision. To determine respondents' writing proficiency before and after the intervention, descriptive statistics, including mean scores, frequencies, and percentages, were computed across the five writing dimensions. These scores were interpreted using a proficiency scale modified from the Common European Framework of Reference for Languages (CEFR) and DepEd Order No. 31, s. 2012 guidelines. To evaluate whether statistically significant changes in

writing proficiency occurred following the AI-assisted revision intervention, the study employed a repeated-measures ANOVA at the 0.05 level of significance. These empirical findings were then synthesized into a structured development plan, ensuring that the proposed intervention program, Project A.I.-V.O.I.C.E., was directly responsive to the specific writing gaps identified in the study.

2.6 Ethical Considerations

The researcher adhered to strict ethical protocols throughout the study. Most importantly, formal Ethics Clearance was obtained from the relevant institutional review board and administrative authorities prior to data collection. All participants were provided with informed consent forms and were fully briefed on the study's purpose and their right to voluntary participation (Creswell & Creswell, 2018). To protect the participants, anonymity and confidentiality were strictly maintained; identifiers were removed from all essay outputs, and data were stored in a secure, restricted-access digital environment.

3. Results

This section presents the results of the pre-test and post-test assessments, using descriptive and inferential statistics to evaluate how AI language tools enhance writing proficiency among non-English-major students. By analyzing five key dimensions—tone and style, contextual understanding, emotional resonance, critical thinking, and originality—this study establishes the empirical basis for a targeted writing intervention program.

Table 1. Level of writing proficiency of the respondents in their written output before AI assistance (N = 77).

Areas		**Before**						
		9-10 (VP)	**7-8 (P)**	**5-6 (E)**	**3-4 (D)**	**0-2 (B)**	**Mean**	**Level**
Tone and Style	F	0	0	51	25	1	4.78	Emerging
	%	0	0	*66.2*	32.5	1.3		
Contextual Understanding	F	0	3	62	11	1	5.39	Emerging
	%	0	3.9	*80.5*	14.3	1.3		
Emotional Resonance	F	0	12	56	8	1	5.62	Emerging
	%	0	15.6	*72.7*	10.4	1.3		
Critical Thinking	F	0	1	54	21	1	4.82	Emerging
	%	0	1.3	*70.1*	27.3	1.3		
Originality and Creativity	F	0	1	59	15	2	4.94	Emerging
	%	0	1.3	*76.6*	19.5	2.6		
Overall	F	0	0	67	9	1	25.55	Emerging
	%	0	0	*87*	11.7	1.3		

Prior to AI assistance, respondents exhibited an overall "Emerging" proficiency level with a mean score of 25.55, where all specific dimensions—Emotional Resonance (5.62), Contextual Understanding (5.39), Originality and Creativity (4.94), Critical Thinking (4.82), and Tone and Style (4.78)—similarly registered at the Emerging level.

Table 2. Level of writing proficiency of the respondents in their written output after AI assistance (N = 77).

Areas		**After**						
		9-10 (VP)	**7-8 (P)**	**5-6 (E)**	**3-4 (D)**	**0-2 (B)**	**Mean**	**Level**
Tone and Style	F	2	19	43	13	0	5.70	Emerging
	%	2.6	24.7	*55.8*	16.9	0		
Contextual Understanding	F	4	31	36	6	0	6.39	Proficient
	%	5.2	40.3	*46.8*	7.8	0		
Emotional Resonance	F	1	41	30	5	0	6.47	Proficient
	%	1.3	*53.2*	39	6.5	0		
Critical Thinking	F	2	20	43	12	0	5.75	Emerging
	%	2.6	26	*55.8*	15.6	0		
Originality and Creativity	F	2	16	49	10	0	5.66	Emerging
	%	2.6	20.8	*63.6*	13	0		
Overall	F	4	31	37	5	0	29.97	Emerging
	%	5.2	40.3	*48.1*	6.5	0		

Following AI integration, overall writing proficiency improved to a mean score of 29.97, placing it at the upper end of the Emerging level and nearly reaching the Proficient category. Post-test results revealed that Emotional Resonance (6.47) and Contextual Understanding (6.39) advanced to the Proficient level, whereas Critical Thinking (5.75), Tone and Style (5.70), and Originality and Creativity (5.66) remained at the Emerging level despite experiencing numerical increases.

Crucially, the study identified four specific writing gaps—Critical Thinking, Tone and Style, Originality and Creativity, and Contextual Understanding—which serve as the primary targets for the proposed intervention program; notably, Contextual Understanding, notwithstanding its proficient mean, was retained for inclusion as a significant portion of the population remained at the Emerging level, necessitating targeted improvement across these four areas.

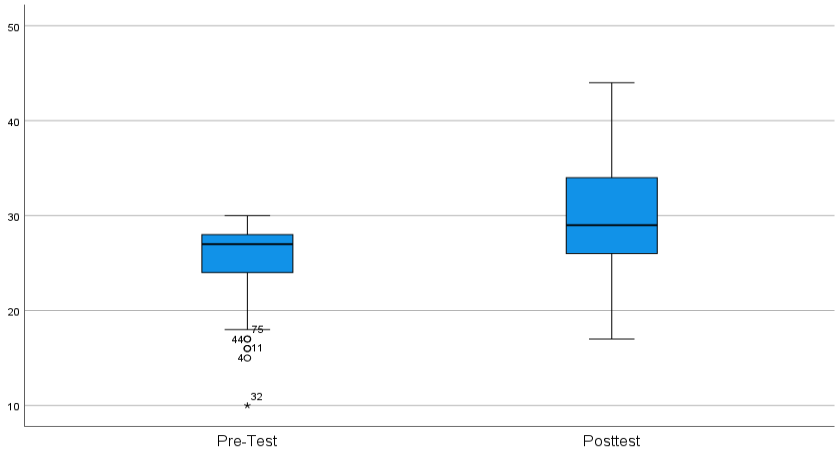

Figure 1. Box plot showing the writing proficiency of the respondents in their written output before and after the AI assistance.

The boxplot in Figure 1 reveals a clear upward shift in the distribution of scores from pre-test to post-test, indicating an overall improvement in writing proficiency. In this visualization, the median post-test score is noticeably higher than the pre-test score, suggesting that typical performance improved after the intervention. Additionally, the post-test box appears wider with longer whiskers, reflecting greater variability in scores; this indicates that while many respondents improved, the degree of improvement varied across individuals. In contrast, the pre-test distribution is more compressed, with several low-end outliers that indicate lower, less dispersed performance initially. The reduction in extremely low scores and the upward shift of the entire distribution in the post-test further emphasize the positive impact of AI assistance on respondents' writing proficiency.

Table 3. Difference between the respondents' writing proficiency before and after the use of AI language tools.

Areas	Mean		Mean Difference (MD=After – Before)	Fc	Sig.	Partial Eta squared
	Before	After				
Tone and Style	4.78	5.70	0.92	33.28*	0.000	.305
Contextual Understanding	5.39	6.39	1.00			
Emotional Resonance	5.62	6.47	0.84			
Critical Thinking	4.82	5.75	0.94			
Originality and Creativity	4.94	5.66	0.73			

Note: *Significant

Table 3 details the significant improvements in writing proficiency following the AI intervention. All five dimensions showed increased mean scores, with the highest gain in Contextual Understanding (MD = 1.00), followed by Critical Thinking (MD = 0.94), Tone and Style (MD = 0.92), Emotional Resonance (MD = 0.84), and Originality and Creativity (MD = 0.73). The computed F-value of 33.28 (p = 0.000) confirms that these differences are statistically significant. Furthermore, a partial eta squared of .305 indicates a moderate to large effect size, accounting for a substantial proportion of the variance in proficiency. Because the p-value (0.000) is well below the 0.05 alpha level, the findings indicate that respondents demonstrated statistically significant improvements in proficiency, as reflected in their written outputs, following AI-assisted revision. However, these findings should be interpreted within the methodological context of the study's purposive sampling procedure and quasi-experimental single-group design.

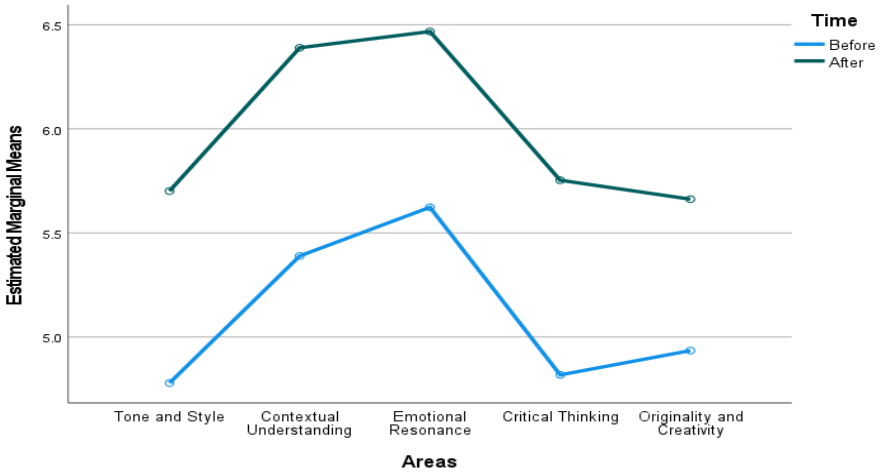

Figure 2. Profile plot describing the difference between the respondents' writing proficiency before and after the use of AI language tools.

Figure 2 illustrates the profile plot of writing proficiency across the five dimensions, showing a clear upward shift following the intervention. The widest gaps appear in Contextual Understanding and Emotional Resonance, suggesting substantial enhancement, while Tone and Style and Critical Thinking show marked upward shifts. Although Originality and Creativity reflect the smallest increase, the parallel upward trend across all areas confirms that AI tools had a uniformly positive effect. This trend suggests that AI integration serves as a comprehensive linguistic bridge rather than a superficial fix. The growth in Contextual Understanding and Emotional Resonance suggests that the tools enable non-English majors to articulate complex sentiments and situational relevance that were previously hindered by vocabulary gaps. Furthermore,

the "ranking up" effect indicates that AI provides a functional "launchpad," allowing students to pivot cognitive energy from mechanical construction to conceptual substance. Ultimately, the plot confirms the cohort's successful transition from a baseline of "Emerging" to "Proficient" in academic standards.

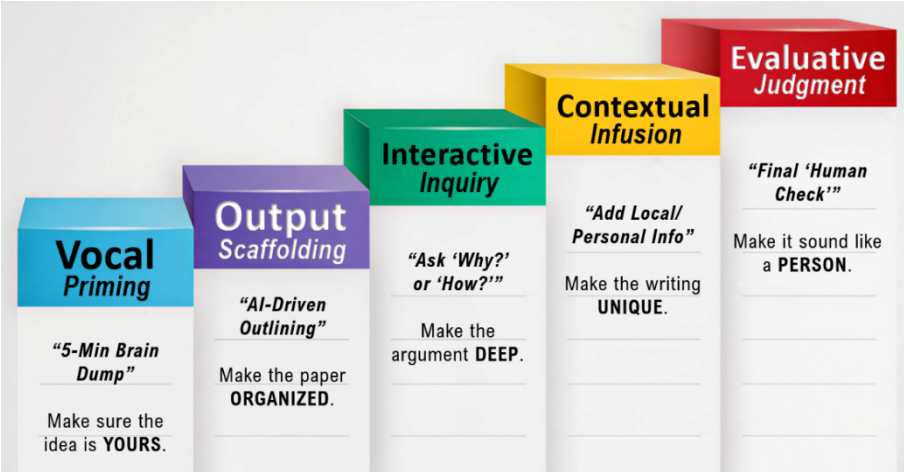

Figure 3. Proposed writing intervention program: Project A.I.-V.O.I.C.E.

Project A.I.-V.O.I.C.E. (Augmented Intelligence: Validating Originality, Insight, and Compositional Excellence) is a proposed writing intervention program developed as a strategic response to the study's finding that while AI tools significantly improved overall writing proficiency (M=25.55 to 29.97) and eliminated "Beginning" level scores, they left a majority of students at an "Emerging" level across four critical dimensions: Originality and Creativity (63.6%), Critical Thinking (55.8%), Tone and Style (55.8%), and Contextual Understanding (46.8%). Grounded in the concept of AI as a "Digital More Knowledgeable Other" (MKO), the program aims to transition students from passive users to "Senior Editors" by targeting these gaps through four key objectives: "humanizing" content to counteract algorithmic homogenization, utilizing AI as a "Socratic Sparring Partner" to deepen logic, cultivating "evaluative judgment" to eliminate robotic prose, and anchoring narratives in authentic professional reality. These goals are operationalized through the five-phase V.O.I.C.E. Framework—consisting of Voice Priming (initial brain dump), Output Scaffolding (logical organization), Interactive Inquiry (critical debate with the AI), Contextual Infusion (localizing content), and Evaluative Judgment (final human-voice audit)—a "human-first" workflow designed to ensure that technology acts as a scaffold for cognitive growth rather than a substitute for independent thought.

4. Discussion

The primary finding of this study is that purposively selected non-English majors with initially low writing proficiency demonstrated statistically significant improvements in their writing proficiency following AI-assisted revision, supporting the study's exploratory investigation into AI-assisted writing support among academically vulnerable learners. Prior to AI assistance, the respondents' baseline proficiency was predominantly at the "Emerging" level, characterized by pronounced difficulties with independent academic writing, rhetorical mastery, and critical synthesis. Following the integration of AI tools, there was a clear upward shift in overall performance, most notably the complete elimination of the lowest "Beginning" tier. Visual evidence from the box plots suggests that AI-assisted revision may have raised the foundational written output performance "floor," minimizing extremely low scores among the selected respondents. Furthermore, profile plot distributions illustrate a uniform upward trajectory, with the most substantial growth occurring in contextual understanding and emotional resonance.

However, a critical analysis of the population frequency distribution reveals a nuanced reality: while aggregate mean scores improved, the vast majority of students still remained anchored at the "Emerging" level across four specific dimensions—critical thinking, originality and creativity, tone and style, and contextual understanding. These findings suggest that AI may serve as a structural scaffold to support students during the revision process; however, it does not automatically confer independent scholarly mastery. Moreover, the observed improvements primarily reflect short-term enhancement in the quality of revised written outputs within the AI-assisted revision context rather than definitive evidence of enduring independent writing proficiency development.

These findings synthesize seamlessly with established sociocultural and linguistic frameworks, particularly Vygotsky's (1978) concept of the Zone of Proximal Development, by positioning AI as a digital "More Knowledgeable Other" that bridges the gap between a student's current limitations and academic expectations. The significant reduction in writing anxiety and the corresponding peak in emotional resonance provide robust empirical support for the Affective Filter Theory (Krashen, 1982), suggesting that AI-assisted revision may help reduce psychological barriers and facilitate a smoother flow of ideas (Zhu et al., 2026). Furthermore, the uniform structural improvements echo international and local studies confirming that generative AI excels at organizational coherence and academic output (Coracero, 2024; Dingal et al., 2024; Escalante et al., 2023; Song & Song, 2023). However, a stark discrepancy emerges regarding higher-order cognitive functions. Consistent with warnings in contemporary literature, the data suggest that students frequently succumbed to "cognitive miserliness," passively adopting machine-generated logic rather than actively challenging it (Mollick & Mollick, 2023). This over-reliance directly accounts for the narrower margins of growth observed in originality and creativity, validating concerns that Large Language Models naturally default to

homogenized, conventional prose that lacks genuine human insight (Espartinez, 2024; Garcia et al., 2025; Li et al., 2023; Russell & Norvig, 2021). Consequently, individual excellence remains tethered not to the AI itself, but to the student's capacity for critical engagement and "evaluative judgment" (Flynn & Allen, 2025).

It is also important to interpret these findings within the methodological context of the study's criterion-based sampling procedure. Because only respondents performing at the "Beginning" to "Emerging" levels were selected for the intervention phase, the possibility of regression-to-the-mean effects cannot be fully excluded. In repeated-measures research, lower initial scorers may naturally show some improvement on subsequent assessments, independent of intervention effects (Barnett et al., 2004). Nevertheless, the observed improvements in this study demonstrated consistent upward trends across all measured writing dimensions. They aligned closely with the existing literature on the benefits of AI-assisted writing support. However, the researcher also recognizes that repeated engagement with the same writing task and revision process may have contributed to some degree of performance improvement independent of AI-assisted support. Thus, while regression to the mean may partially account for score changes, the findings still provide meaningful exploratory evidence of the potential instructional value of AI-assisted revision among academically vulnerable writers.

The practical implications of these findings highlight the need for a strategic shift in higher education pedagogy, transitioning students from passive AI consumers to active "Senior Editors." Because the frequency distributions clearly indicate that a majority of the population still requires targeted intervention in four specific writing dimensions, this study proposes the theoretically grounded Project A.I.-V.O.I.C.E. (Augmented Intelligence: Validating Originality, Insight, and Compositional Excellence) framework as a potential pedagogical model for future AI-assisted writing interventions. This framework introduces a five-phase workflow—Voice Priming, Output Scaffolding, Interactive Inquiry, Contextual Infusion, and Evaluative Judgment—designed to redefine AI as a Socratic sparring partner rather than an automated substitute for intellectual labor. It is important to note, however, that the framework itself was not directly implemented or empirically tested within the present study and therefore remains a theoretically derived intervention model requiring future validation. By forcing students to manually graft local, dimension-specific artifacts into their drafts and rigorously challenge the AI's logical structures, the program anchors academic writing in authentic professional reality, embedding them firmly within their discourse community. This real-world application ensures that students retain executive control over the composition process, effectively counteracting algorithmic homogenization (Russell & Norvig, 2021) and preserving their unique professional identity.

Despite the robust insights generated, it is imperative to acknowledge the limitations that frame this study. The research was conducted with a specific sample of 77 non-English majors, which inherently limits the generalizability of the findings to other diverse academic disciplines or distinct linguistic demographics. Another

methodological limitation stems from the use of a single-group pretest-posttest design without a non-AI revision control group. Because of this, improvements in post-test performance may have been influenced not only by AI-assisted feedback but also by factors such as extended revision time, repeated exposure to the writing task, familiarity with the analytic rubric, and the natural benefits of revising an earlier draft. Consequently, the study cannot fully isolate the independent effect of AI language tools from the general benefits associated with revision practice. Future research employing randomized or comparative control-group designs is therefore recommended to strengthen causal inference.

A further limitation arises from the purposive selection of respondents who initially performed at lower proficiency levels. Because participants were selected based on "Beginning" to "Emerging" pretest scores, the study may be susceptible to regression-to-the-mean effects, in which lower scorers may naturally show higher scores on subsequent testing, independent of intervention influence. Consequently, the findings should be interpreted cautiously and not generalized as definitive proof of AI effectiveness across broader student populations. Future studies employing randomized sampling procedures and comparison groups are therefore strongly recommended.

Additionally, reliance on current generative AI platforms introduces a degree of technological dependency, as these proprietary algorithms undergo continuous updates, and the exact nature of their structural outputs and inherent stylistic biases will inevitably shift. Furthermore, while the methodology effectively captured short-term improvements in the proficiency reflected in respondents' revised written outputs within immediate pre-test and post-test environments, it did not measure long-term cognitive retention or determine whether these observed gains would persist independently once AI-assisted support is removed over an extended period. Finally, evaluating nuanced dimensions such as originality and tone introduces an unavoidable degree of subjective interpretation, potentially biasing assessments despite the use of standardized rubrics. Moreover, because the post-test outputs involved revised versions of respondents' earlier essays, the study acknowledges the possibility that evaluator familiarity may influence aspects of the scoring process, despite adviser moderation and rubric standardization. It must also be acknowledged that although the study proposes the Project A.I.-V.O.I.C.E. framework as a theoretically informed instructional model derived from the findings, the framework itself was not directly implemented or experimentally evaluated within the present research design.

Future inquiry must prioritize longitudinal research to investigate whether the cognitive scaffolding provided by AI language tools translates into enduring, unassisted writing proficiency over an extended academic timeline. Future empirical investigations are strongly recommended to evaluate the instructional effectiveness, feasibility, and long-term applicability of Project A.I.-V.O.I.C.E. across varying institutional contexts and broader student populations. Researchers should also explore the psychological dynamics of human-AI collaboration, specifically examining how continuous interaction with generative tools impacts a student's

authorial confidence and perceived academic integrity. Ultimately, this study concludes that while Artificial Intelligence may serve as a transformative digital mediator that supports improvements in foundational written output quality in AI-assisted revision contexts, it is not a panacea for academic literacy. True compositional excellence remains an exclusively human endeavor, requiring educational institutions to actively cultivate students' evaluative judgment so that technology amplifies rather than replaces human intellect (Flynn & Allen, 2025).

5. Conclusion

The primary inquiry of this research sought to determine the efficacy of AI language tools in enhancing the writing proficiency of non-English majors and to identify whether such integration could bridge specific gaps in scholarly composition. The empirical data reveal that purposively selected non-English major students with initially low writing proficiency demonstrated significant improvements in their writing proficiency, as reflected in their revised written outputs following AI-assisted revision. These findings must nevertheless be interpreted within the methodological limitations of the study's quasi-experimental single-group design and criterion-based sampling procedure. Overall proficiency shifted from an initial mean of 25.55 to the upper threshold of the "Emerging" level at 29.97, which approached the "Proficient" category. However, the findings demonstrate that AI integration does not serve as a total substitute for human cognitive depth. Importantly, these findings should be interpreted as evidence of short-term enhancement within AI-assisted revision tasks rather than conclusive proof of enduring independent writing proficiency development.

While a statistically significant improvement was observed across all dimensions ($F = 33.28$, $p < .05$), higher-order skills, including Originality and Creativity, Critical Thinking, and Tone and Style, remain strictly within the "Emerging" category. Furthermore, although the mean for Contextual Understanding reached the "Proficient" level, it remains a critical target for intervention as the majority of the student population continues to perform at the "Emerging" level in this domain. Thus, the study achieves its objective by suggesting that AI-assisted revision may serve as a useful structural scaffold while still requiring substantial human cognitive engagement for higher-order writing skills, particularly in refining originality, critical thinking, and authorial voice.

The implications of these findings may offer meaningful insights for both contemporary pedagogy and the broader academic field. By pinpointing the specific limitations of AI—namely, its tendency toward "algorithmic homogenization"—this work contributes a nuanced understanding of human-AI collaboration to the existing body of knowledge. It argues that the value of AI in the classroom lies not in automating text, but in its role as a diagnostic tool that highlights where human intervention is most needed. For the industry and academic institutions, this research provides a practical roadmap for evolving writing curricula from traditional drafting

to evaluative editing. It establishes that the study's relevance lies in its move away from the binary debate over "using versus banning" AI, focusing instead on how students can maintain agency and scholarly excellence in an automated era.

To translate these findings into practice, this study proposes the Project A.I.-V.O.I.C.E. framework as a theoretically grounded intervention model that may guide future AI-assisted writing instruction and pedagogical innovation. Because the framework itself was not empirically implemented or evaluated within the present investigation, future studies are strongly recommended to examine its practical effectiveness across broader instructional contexts. This five-phase intervention is designed to target the identified gaps in evaluative judgment and original composition. Practitioners should specifically redesign assessment rubrics to reward the "Voice Priming" and "Interactive Inquiry" stages, ensuring students remain the primary architects of their work.

Furthermore, future research should pursue longitudinal studies to examine how consistent use of this intervention influences long-term writing autonomy across disciplines. Ultimately, this study underscores that while technology may provide the tools for expression, enduring mastery of the written word remains a deeply human endeavor, necessitating a critical partnership between intelligence and insight.

Acknowledgment

Sincere gratitude is extended to all individuals and institutions whose technical, administrative, and moral support contributed to the completion of this study. Special appreciation is given to Pangasinan State University, School of Advanced Studies; the researcher's home institution; and Urdaneta City University, the primary study site, for the permissions, cooperation, and resources provided. Profound appreciation is extended to Dr. Nhorly U. Domenden, Dean of the College of Human Sciences and CEO/Founder of the Wundt Psychological Institute, for professional mentorship, technical guidance, and unwavering support.

Gratitude is likewise extended to the Thesis Adviser for intellectual guidance and mentorship; to the Thesis Panel Members for their constructive evaluation; and to the Critic Reader and Statistician for linguistic refinement, data interpretation, and statistical support. The panel of expert evaluators is also acknowledged for validating the research instruments and strengthening the reliability and academic integrity of the study. Heartfelt thanks are given to the student-respondents for their participation, and to family, friends, and a dedicated personal partner for the encouragement, prayers, logistical assistance, and material support provided.

No external funding from government agencies, academic grants, or private foundations was received. The study was self-funded.

Conflict of Interest Statement

The author declares no conflict of interest.

References

Ahmed, A. (2024, September 10). ChatGPT tops AI tools among students, 86% use AI for studies: Survey. *Digital Information World*. Retrieved from https://www.digitalinformationworld.com/2024/09/chatgpt-tops-ai-tools-among-students-86.html

Archana, S. N., Renjith, V. R., Padmakumar, P. K., C, S., & Aboobaker, N. (2025). AI assisted learning and research: An exploratory study among university students and scholars. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00814-x.

Barnett, A. G., van der Pols, J. C., Dobson, A. J. (2004). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299.

Coracero, R. M. M. (2024). The utilization of artificial intelligence (AI) writing assistants to the writing proficiency. *International Journal of Research Publications*, *50*(1). https://doi.org/10.47119/ijrp1001501620246659.

Council of Europe. (n.d.). Common European framework of reference for languages (CEFR): Level descriptions. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Incorporated.https://books.google.com/books?hl=en&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=Creswell,+J.+W.,+%26+Creswell,+J.+D.+(2018).+Research+design:+Qualitative,+quantitative,+and+mixed++%09methods+approaches.+SAGE+Publications,+Incorporated.+&ots=YEAPKQsstH&sig=lpixwKxlMcuuxg7Y37yBiUFJEqY

Department of Education. (2012). DepEd order no. 31, s. 2012: Policy guidelines on the implementation of the grades 1 to 10 of the K to 12 basic education curriculum (BEC) effective school year 2012-2013. Retrieved from https://www.scribd.com/document/231788446/DO-No-31-s-2012

Dingal, S. M. L., Balbin, C. J. P., Buenafe, S. G. L., Buenavista, P. E. M., Cabrera, C. A., Lorica, V. M. T., Olalo, A. M., Pataganao, L. a. A., Wahing, M. R. A., Clamares, K. J. M., & Pelandas, A. M. O. (2024). Artificial intelligence (AI) usage and its influence to the students' academic writing: A quantitative – correlation investigation. *International Journal of Research and Innovation in Social Science*, *VIII*(IV), 1621–1627. https://doi.org/10.47772/ijriss.2024.804216.

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00425-2.

Espartinez, A. S. (2024). Exploring student and teacher perceptions of ChatGPT use in higher education: A Q-Methodology study. *Computers and Education Artificial Intelligence, 7, 100264.* https://doi.org/10.1016/j.caeai.2024.100264.

Flynn, L. E., & Allen, L. K. (2025). Disentangling the influences of creativity and verbal fluency on writing. *Journal of Learning Analytics, 12*(2), 67–82. https://doi.org/10.18608/jla.2025.8857.

Fraenkel, J., Hyun, H., & Wallen, N. (2011). *How to design and evaluate research in education*. McGraw-Hill Education. https://studentebookhub.com/wpcontent/uploads/2024/preview/9781259913839.pdf

Freeman, J. (2025). Student generative AI survey 2025. *Higher Education Policy Institute.* Retrieved from https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/

Garcia, N. P. D., Maghanoy, N. D. R. G., Ahmad, N. F. S. H., Manabat, N. a. D., & Arnoco, N. J. P. P. (2025). Chatgpt, Grammarly, and Quillbot: Perceptions of students and teachers towards the use of AI tools in writing. *Journal of English Language Teaching and Applied Linguistics*, *7*(3), 161–167. https://doi.org/10.32996/jeltal.2025.7.3.16.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Pergamon Press. https://www.academia.edu/download/56454923/Communicative_Approach_by_Stephen_Crashen.pdf

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively?. *Computers and Education Artificial Intelligence*, *4*, 100140. https://doi.org/10.1016/j.caeai.2023.100140.

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00602-7.

Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. *Social Science Research Network (SSRN).* https://doi.org/10.2139/ssrn.4475995.

Russell, S. J., & Norvig, P. (2021). Artificial Intelligence: A modern approach (Pearson series in Artificial Intelligence). *Amazon.com: Books.* Retrieved from https://www.amazon.com/Artificial-Intelligence-A-Modern-Approach/dp/0134610997

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin. https://www.tandfonline.com/doi/pdf/10.1198/jasa.2005.s22

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://www.jstor.org/stable/j.ctvjf9vz4

Zhu, Y., Zhang, Y., & Yao, C. (2026). Navigating emotional resonance and language identity through world Englishes education: An experimental intervention with EFL learners in China. *Frontiers in Psychology, 17,* 1793431. https://doi.org/10.3389/fpsyg.2026.1793431

Author Contributions: Garcia, M.; Conceptualization, methodology, and investigation; data curation, formal analysis, and manuscript writing; data visualization, critical review and editing, funding acquisition, resource provision, and supervision.