

ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 05 No. 01. Jan-March 2026.Page#.8687-8696
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.23954410

Generative versus Rule-Based AI in ESL Academic Writing: ChatGPT's Dominance and Its Concentration at the Brainstorming Stage

Maria Bilal
MPhil Scholar, Department of Applied Linguistics, Government College University Faisalabad, Pakistan
mariakhan275@icloud.com
Shahida Naz*
Assistant Professor, Department of Applied Linguistics, Government College University Faisalabad, Pakistan
**Corresponding author shahidanaz@gcuf.edu.pk*

Abstract
This study examines why generative AI has displaced rule-based Automated Writing Evaluation (AWE) tools as the primary AI writing resource among Pakistani ESL undergraduates, and what this shift implies for the specific writing sub-skills such tools appear to support. Survey data from 300 BS English students at a public-sector Pakistani university show that ChatGPT accounted for 84.0% of primary AI-tool usage, decisively outpacing Grammarly (10.0%), QuillBot (4.3%), and Writefull (1.7%) a marked reversal of the Grammarly-dominant usage patterns reported in earlier ESL literature. Usage-stage data further show that AI support was concentrated overwhelmingly at the brainstorming/idea-generation stage of writing (49.7%), well ahead of drafting (19.7%), editing/proofreading (21.0%), and finalization (9.7%). Consistent with this generative, pre-writing usage profile, learners rated Organization and Structure as the writing dimension most positively influenced by AI (perception M = 3.69), inverting the grammar-first hierarchy typical of Grammarly-centred research. However, independent essay performance from a 65-participant AI-free writing subsample showed that this organizational advantage was the only dimension where perception and performance meaningfully converged (essay M = 2.83, the narrowest perception-performance gap of all five dimensions at 0.86 points); grammar, coherence, and academic style dimensions less directly served by ChatGPT's generative strengths showed substantially larger gaps. These findings indicate that ChatGPT's dominance is reshaping which writing sub-skills AI tools are perceived to develop, with genuine but qualified evidence that its brainstorming-stage scaffolding partially transfers to independent structural competence, while its influence on grammar, coherence, and register remains largely confined to assisted-task quality. The study offers a tool-specific account of generative AI's distinctive pedagogical footprint in ESL writing instruction, with implications for how brainstorming-stage AI use should be pedagogically framed.

Keywords: *ChatGPT; generative AI; brainstorming; automated writing evaluation; ESL academic writing; organization and structure; Pakistani higher education*

1. Introduction

Automated Writing Evaluation (AWE) tools such as Grammarly have dominated both ESL classroom practice and the empirical literature on AI-assisted writing for over a decade, valued primarily for surface-level error correction (Cavaleri & Dianati, 2016; O'Neill & Russell, 2019; Karyuatry & Rizqan, 2018). The emergence of generative large language models, and ChatGPT in particular, represents a qualitatively different category of tool: rather than identifying errors in existing text, ChatGPT produces original prose, responds to metalinguistic queries, and engages in extended pedagogical dialogue (Kasneci et al., 2023). This is not simply a new entrant into an existing tool category but a shift in what AI-assisted writing support actually consists of from corrective feedback on a learner's own text to generative co-construction of ideas and structure before the learner has written anything independently.

This shift has outpaced the empirical literature. Much of the foundational research on AI and ESL writing was conducted when Grammarly-type AWE tools were the dominant technology, and consistently found their benefits concentrated in grammatical accuracy, with markedly smaller effects on coherence and organization (Huang et al., 2020; Nami, 2023). Whether this grammar-first benefit hierarchy still holds once ChatGPT becomes the dominant tool and whether ChatGPT's generative strengths produce a different, planning-centred benefit hierarchy remains an open, empirically under-examined question, particularly in Pakistani higher education, where large-scale usage-pattern data linking specific tool choice to specific writing-stage use and dimension-level outcomes are scarce.

This study addresses that question directly, using tool-preference data, writing-stage data, and dimension-level perception and performance data from the same 300-participant (with 65-participant essay subsample) Pakistani undergraduate population to characterize ChatGPT's distinctive footprint in ESL academic writing development.

1.1 Research Objectives

This study aims to: (1) document the extent and pattern of ChatGPT's dominance over rule-based AI writing tools among Pakistani ESL undergraduates

(2) Examine the concentration of ChatGPT use at the brainstorming/pre-writing stage relative to other writing stages; and

(3) evaluate whether this generative, pre-writing usage pattern is associated with a distinctive, organization-centred pattern of perceived and independently demonstrated writing benefit.

1.2 Research Questions

RQ1: To what extent has ChatGPT displaced rule-based AWE tools (Grammarly, QuillBot, Writefull) as the primary AI writing resource among ESL undergraduates?

RQ2: At which stage of the writing process is ChatGPT use concentrated, and what does this suggest about its functional role?

RQ3: Is ChatGPT's dominance and brainstorming-stage concentration associated with a distinctive benefit profile specifically, stronger perceived and demonstrated gains in Organization and Structure relative to other writing dimensions?

1.3 Significance of the Study

Most existing AI-and-ESL-writing research either predates ChatGPT's widespread adoption or treats 'AI writing tools' as a single undifferentiated category. By isolating ChatGPT's specific usage pattern and testing whether it produces a specific, theoretically predictable benefit profile, this study offers instructors and curriculum designers tool-specific rather than generic guidance, and contributes original evidence to the small but growing literature comparing generative and rule-based AI tools within a single ESL population (Mahapatra, 2024; Al-Sharqi & Hashmi, 2023).

2. Literature Review

2.1 Rule-Based AWE Tools: Grammarly, QuillBot, and Writefull

Grammarly, the most widely used AWE system globally, uses natural language processing to flag grammatical errors, spelling mistakes, and stylistic weaknesses (Cavaleri & Dianati, 2016). Meta-analytic evidence shows a moderate-to-large effect of AWE feedback on grammatical accuracy (d = 0.62), but a substantially smaller effect on coherence and organization (d = 0.31) (Huang et al., 2020) the asymmetry this study directly re-examines under ChatGPT-dominant usage conditions. QuillBot, a paraphrasing tool, is valued for increasing sentence variety and register flexibility (Thohir, 2023), while Writefull, trained specifically on academic texts, has shown measurable benefit for academic collocational accuracy (Koltovskaia, 2020). None of these tools is generative in the sense of producing original extended prose or structural outlines from a prompt.

2.2 ChatGPT as a Generative Alternative

ChatGPT's capacity to generate essay outlines, model paragraph transitions, and engage in extended dialogue about structure and argument positions it as functionally distinct from AWE tools (Kasneci et al., 2023; Mahapatra, 2024). Pakistani learners report using ChatGPT primarily for generating essay outlines, improving transitions, and seeking grammatical explanations (Khan & Ghani, 2024). Empirically, Mahapatra (2024) found that participants who used ChatGPT specifically for organizational feedback showed significantly greater improvement in holistic essay quality on organizational subscales, and Al-Sharqi and Hashmi (2023) found ChatGPT-assisted brainstorming produced more developed, multi-perspective essays than independent brainstorming. Kasneci et al. (2023) identify discourse-level organization and argumentation as the areas where ChatGPT's pedagogical potential most clearly exceeds that of AWE tools directly motivating this study's hypothesis that ChatGPT-dominant samples should show an organization-first, rather than grammar-first, benefit hierarchy.

2.3 Writing-Stage Effects and Scaffolding Theory

From a Vygotskian (1978) scaffolding perspective, where in the writing process a tool is used matters as much as which tool is used. Scaffolding applied at the brainstorming stage before any independent cognitive engagement carries a theoretically distinct risk and opportunity profile compared to scaffolding applied at the editing stage, after a learner has already produced independent text. AWE tools, by design, operate almost exclusively at the editing/proofreading stage, correcting text the learner has already produced. ChatGPT's generative capabilities, by contrast, are most naturally deployed before writing begins — a structural difference in tool design that plausibly explains divergent stage-of-use patterns and, this study argues, divergent dimension-level benefit profiles.

2.4 Research Gap

While the benefits of AWE tools for grammar and the generative strengths of ChatGPT for structure are separately documented, no Pakistani study has yet combined tool-preference data, writing-stage data, and dimension-level perception-performance comparison within a single sample to test whether ChatGPT's dominance produces a measurably different and only partially transferable benefit hierarchy than the grammar-first pattern established in the AWE-dominant literature. This is the gap the present study addresses.

3. Methodology

3.1 Design

A quantitative, descriptive-correlational design was employed (Creswell & Creswell, 2018). Tool-preference and writing-stage data were collected from all 300 questionnaire respondents; dimension-level perception data were collected from the 264 respondents reporting active AI tool use; and independent essay-performance data were collected from a 65-participant subsample completing a supervised, AI-free essay task, enabling a direct test of whether ChatGPT's dominant, brainstorming-concentrated usage pattern corresponds to a distinctive and at least partially durable benefit profile.

3.2 Participants and Instruments

Participants were BS English undergraduates at Government College University Faisalabad, recruited through convenience sampling (Dörnyei, 2007). Tool preference was measured through a single-choice item identifying participants' primary AI writing tool (ChatGPT, Grammarly, QuillBot, or Writefull). Writing-stage use was measured through a single-choice item identifying the stage (brainstorming, drafting, editing/proofreading, or finalizing) at which AI support was most frequently used. Dimension-level perception was measured via a 25-item, five-point Likert scale spanning five writing dimensions (α = .87). Independent writing performance was assessed via a supervised, 40-minute, AI-free argumentative essay, scored by a blinded evaluator using a 25-mark, five-dimension rubric.

3.3 Data Analysis

Tool-preference and writing-stage data were analyzed descriptively. Dimension-level perception means and essay performance means were compared to test whether the dimension most associated with ChatGPT's generative affordances (Organization and Structure) showed a distinctive convergence between perception and independently demonstrated performance, relative to the other four dimensions.

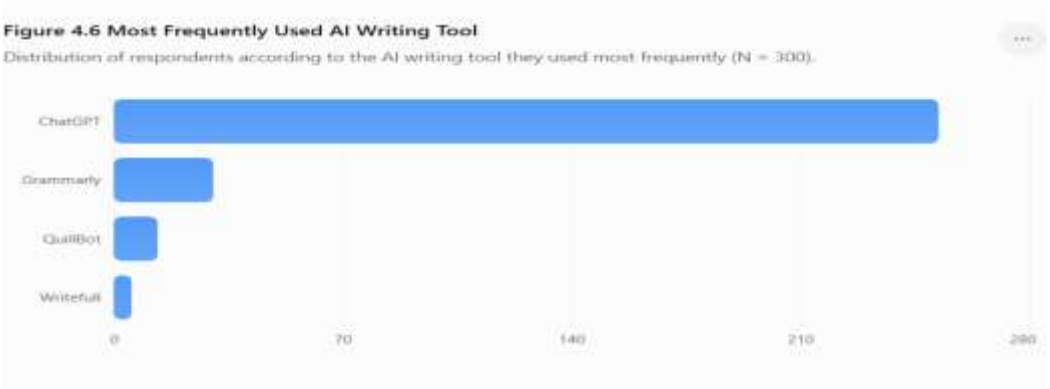
4. Results

4.1 ChatGPT's Dominance over Rule-Based AI Tools

ChatGPT was the primary AI writing tool for 84.0% of the 300-participant sample, decisively outpacing all three rule-based AWE alternatives combined (16.0%).

Table 1. Primary AI Writing Tool Preference (N = 300)

AI Tool	Tool Type	Frequency	Percentage (%)
ChatGPT	Generative (LLM)	252	84.0
Grammarly	Rule-based AWE	30	10.0
QuillBot	Rule-based (paraphrasing)	13	4.3
Writefull	Rule-based AWE (academic)	5	1.7



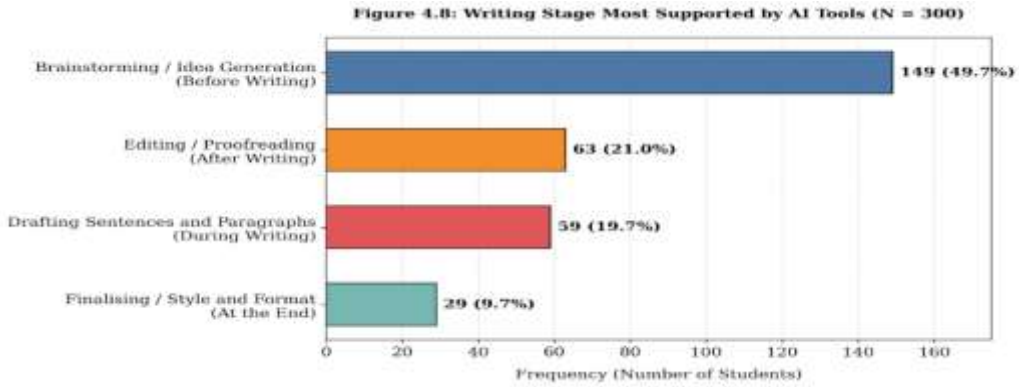
This distribution represents a marked reversal of the Grammarly-dominant usage patterns reported in earlier ESL research (Cavaleri & Dianati, 2016; O'Neill & Russell, 2019), confirming that generative AI has become the primary AI writing resource for this population rather than one option among several comparable tools.

4.2 Concentration of AI Use at the Brainstorming Stage

Writing-stage data show AI support concentrated overwhelmingly at brainstorming/idea-generation, used by nearly half the sample well ahead of any other stage.

Table 2. Writing Stage Most Supported by AI Tools (N = 300)

Writing Stage	Frequency	Percentage (%)
Brainstorming / Idea Generation (before writing)	149	49.7
Editing / Proofreading (after writing)	63	21.0
Drafting Sentences and Paragraphs (during writing)	59	19.7
Finalizing / Style and Format (at the end)	29	9.7



Given that 84% of the sample primarily used ChatGPT, this brainstorming-stage concentration is best interpreted as a direct behavioral signature of generative AI use: ChatGPT is functionally suited to producing outlines and structural suggestions before independent writing begins, in a way that rule-based AWE tools which operate on already-written text structurally cannot replicate.

4.3 Dimension-Level Perception: An Organization-First Hierarchy

Consistent with ChatGPT's generative, pre-writing usage profile, learners rated Organization and Structure as the dimension most positively influenced by AI tools, ahead of Coherence, Academic Style, Grammar, and Vocabulary an inversion of the grammar-first hierarchy typical of Grammarly-dominant research (Aladini et al., 2024; O'Neill & Russell, 2019).

Table 3. Learner Perception Scores by Writing Dimension (n = 264)

Writing Dimension	Perception Mean (/5)	Rank
Organization and Structure	3.69	1st
Coherence and Logical Flow	3.65	2nd
Academic Style and Tone	3.64	3rd
Grammar and Accuracy	3.59	4th
Vocabulary and Lexical Choice	3.56	5th

4.4 Does the Organizational Advantage Hold Under Independent Testing?

The critical test of whether ChatGPT's brainstorming-stage scaffolding produces genuine, transferable benefit rather than merely a favorable impression is whether the organizational advantage seen in perception data also appears in independently assessed, AI-free essay performance. It does, though only partially.

Table 4. Perception–Performance Comparison, Organization and Structure versus Other Dimensions

Writing Dimension	Perception M	Essay M (N=65)	Gap	Interpretation
Organization and Structure	3.69	2.83	0.86	Narrowest gap; only dimension where perception rank matches performance rank
Vocabulary and Lexical Choice	3.56	2.69	0.87	Partial alignment
Grammar and Accuracy	3.59	2.49	1.10	Moderate discrepancy
Coherence and Logical Flow	3.65	2.55	1.10	Noticeable discrepancy
Academic Style and Tone	3.64	2.46	1.18	Largest discrepancy

Organization and Structure achieved both the highest perception mean and the highest independent essay-performance mean (M = 2.83/5) of all five dimensions, and the narrowest perception-performance gap (0.86 points) the only dimension in the entire study where perception rank and performance rank converge. By contrast, Academic Style and Tone, a dimension far less directly served by ChatGPT's generative structural strengths, showed the largest gap (1.18 points) despite a similarly positive perception score (M = 3.64).

5. Discussion

These findings directly answer the study's three research questions. First, ChatGPT has not merely gained a foothold alongside AWE tools but has become the overwhelmingly dominant AI writing resource for this population (84.0%), fundamentally changing what 'AI-assisted writing' means in practice for Pakistani ESL undergraduates (RQ1). Second, this generative tool's use is heavily concentrated at the brainstorming stage (49.7%), a pattern structurally consistent with ChatGPT's outline-generation and idea-development capabilities and inconsistent with how rule-based AWE tools, which operate on already-written text, are typically used (RQ2). Third, this brainstorming-stage concentration corresponds to a distinctive, organization-first benefit hierarchy that is not merely a perceptual artefact: it is the one dimension where independent essay performance corroborates the perception data, providing genuine, if qualified, evidence that ChatGPT-supported structural planning partially transfers to independent writing (RQ3).

This pattern is theoretically coherent. From a Vygotskian (1978) perspective, ChatGPT's brainstorming-stage scaffolding operates on the generative, ideational component of writing arguably a more transferable cognitive skill (learning to structure an argument) than the highly rule-governed, convention-bound skills of grammar and register, which require sustained, deliberate practice to internalize (Lantolf & Thorne, 2006). This may explain why Organization and Structure shows partial transfer while Academic Style and Tone, one of the most culturally embedded and difficult-to-internalize

dimensions of academic writing (Lantolf & Thorne, 2006), does not, despite comparably positive perceptions of both.

At the same time, the 0.86-point gap for Organization and Structure, while the narrowest in the study, is not trivial: even the dimension most favorably served by ChatGPT's generative capabilities does not fully transfer to independent performance. This tempers any conclusion that generative AI 'solves' the organizational weaknesses long documented in ESL writing research; rather, it suggests generative AI shifts the locus of AI benefit from surface accuracy toward structural planning, with transfer to independent competence still partial rather than complete consistent with Liu's (2023) broader conclusion that AI-tool satisfaction is not reliably matched by measurable independent performance gains.

6. Conclusion

This study demonstrates that ChatGPT's rise to near-total dominance (84.0%) over rule-based AWE tools among Pakistani ESL undergraduates is accompanied by a distinctive, theoretically coherent behavioral and pedagogical signature: heavy concentration of AI use at the brainstorming stage, and a corresponding organization-first rather than grammar-first benefit hierarchy. This hierarchy is not merely perceptual it is the only dimension in which independently assessed writing performance corroborates learners' perceptions, offering genuine evidence that ChatGPT's generative strengths provide structural, planning-stage scaffolding that partially transfers to independent writing, even as its influence on grammar, coherence, and academic register remains largely confined to assisted-task quality.

7. Implications

7.1 Pedagogical Implications

Instructors should explicitly leverage ChatGPT's demonstrated organizational strength by structuring brainstorming-stage AI use around scaffolded, learner-driven outlining tasks rather than passive acceptance of AI-generated structures, while recognizing that grammar, coherence, and academic register still require dedicated, AI-independent instructional attention, as generative AI use alone does not appear to close these gaps.

7.2 Tool-Selection Implications

Because different tool types show different, only partially overlapping benefit profiles generative AI for structural planning, AWE tools for surface accuracy institutions and instructors should consider deliberate multi-tool pedagogical guidance rather than treating 'AI writing tools' as an undifferentiated category, consistent with Rahman and Watanabe's (2022) suggestion that multi-tool use may produce more comprehensive writing development than reliance on any single tool.

7.3 Theoretical Contribution

This study extends Vygotskian scaffolding theory into tool-specific territory, showing that the type of AI tool and the writing stage at which it is deployed jointly shape which writing competencies show evidence of transfer a refinement of prior research that has largely treated AI writing tools as functionally interchangeable.

8. Limitations

The comparison between ChatGPT and rule-based tools is necessarily unbalanced, as only 16% of the sample primarily used a non-ChatGPT tool, limiting statistical power for direct tool-by-tool comparison. The cross-sectional design cannot establish whether the organizational advantage reflects genuine skill transfer or reproduction of AI-modelled templates. Findings are drawn from a single institution using convenience sampling, and the essay subsample (n = 65) was self-selected, which may affect the generalizability of the performance findings.

9. Recommendations and Future Research

Future research should employ controlled, experimental designs directly comparing ChatGPT-assisted and AWE-assisted brainstorming on subsequent independent writing performance, examine whether explicit instruction in critically evaluating AI-generated outlines strengthens the transfer of organizational competence, and extend this tool-specific comparative approach to other generative AI models as they become more widely adopted in ESL contexts.

References

Al-Sharqi, L., & Hashmi, K. (2023). The influence of ChatGPT on EFL learners' brainstorming and essay writing quality. Arab World English Journal, 14(3), 56–71.

Aladini, A., Rehman, M., & Shahzad, K. (2024). Attitudes of Pakistani undergraduate ESL students toward artificial intelligence in improving English writing skills. Applied Linguistics Research Journal, 7(3), 112–129.

Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. Journal of Academic Language and Learning, 10(1), A223–A236.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE.

Dörnyei, Z. (2007). Research methods in applied linguistics. Oxford University Press.

Huang, Y., Liang, H., & He, X. (2020). Effects of automated writing evaluation on EFL student writing quality: A meta-analysis. CALICO Journal, 37(2), 141–164.

Karyuatry, L., & Rizqan, M. D. (2018). Grammarly as a tool to improve students' writing quality. JSSH (Jurnal Sains Sosial dan Humaniora), 2(1), 83–89.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274.

Khan, M. A., & Ghani, M. (2024). Investigating the role of AI writing tools in developing academic writing skills: Learner perceptions at a Pakistani university. Journal of Applied Linguistics and TESOL, 6(1), 34–51.

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly and its potential for developing writing proficiency. Assessing Writing, 44, Article 100468.

Lantolf, J. P., & Thorne, S. L. (2006). Sociocultural theory and the genesis of second language development. Oxford University Press.

Liu, Q. (2023). AI tools in L2 writing development: A systematic review of empirical studies. Language Teaching Research. Advance online publication.

Mahapatra, S. (2024). Exploring the role of ChatGPT in EFL writing: Learner experiences and writing quality in a Bangladeshi tertiary context. Journal of English for Academic Purposes, 68, Article 101345.

Nami, F. (2023). Exploring EFL learners' perceptions of AI writing assistants: Benefits, limitations, and implications for writing development. Computers & Education, 198, Article 104764.

O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback provided by Grammarly. Australasian Journal of Educational Technology, 35(1), 42–56.

Rahman, H., & Watanabe, Y. (2022). ESL learners' perceptions of AI writing assistance: A multi-national comparative survey. Asian EFL Journal, 24(2), 78–102.

Thohir, L. (2023). QuillBot as an AI writing tool: EFL learners' perceptions of its use in academic writing tasks. Journal of English Language Teaching and Linguistics, 8(1), 89–104.

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