

The Quality of Scientific Papers of Undergraduate Students in Regional Language and Literature Studies Lectures Before and After Using ChatGPT as a Writing Tool

Ari Wulandari¹, Arif Budiman², Tri Ratna Herawati³, Primasari Wahyuni⁴, Diaz Danu Setyawan⁵

Universitas PGRI Yogyakarta, Indonesia^{1,3,4,5}

Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia²

Email Correspondence: kinoyasan@gmail.com

Abstract

Background:

This study addresses the growing use of ChatGPT in academic writing. Grounded in constructivism, the study investigates changes in the quality of scientific papers produced by undergraduate students in the Regional Language and Literature Studies course at University “A” before and after the use of ChatGPT as a writing support tool.

Methodology:

A quasi-experimental pretest–posttest design using document analysis was applied. The data consisted of 30 student papers written before and after ChatGPT use. Assessment covered five aspects: ideas, organization, grammar, diction, and spelling.

Findings:

The findings indicate substantial improvements in students’ scientific papers after the use of ChatGPT. The mean writing score increased from 77.9 before ChatGPT-assisted revision to 89.0 after revision, and 27 of the 30 papers improved to the excellent category. The strongest improvements were observed in organization, grammar, diction, and spelling, whereas gains in idea development were comparatively more modest.

Conclusion:

The findings suggest that ChatGPT-assisted revision can enhance the technical quality of students’ scientific writing. However, lecturer guidance remains necessary to support critical thinking, independent analysis, and the responsible use of generative AI in academic writing.

Originality:

This study contributes empirical evidence on changes in the quality of students’ scientific writing associated with ChatGPT-assisted revision within the context an undergraduate language and literature course.

Keywords : ChatGPT; Scientific Papers; Undergraduate Students; Regional Language and Literature Studies; Quality of Academic Writing

DOI	: 10.24903/sj.v11i2.2353
Received	: February 2026
Accepted	: July 2026
Published	: August 2026
How to cite this article (APA)	Wulandari, A., Budiman, A., Herawati, T. R., Wahyuni, P., & Setyawan, D. D. (2026). The quality of scientific papers of undergraduate students in regional language and literature studies lectures before and after using ChatGPT as a writing tool. <i>Script Journal: Journal of Linguistics and English Teaching</i> , 11(2), 340–417. https://doi.org/10.24903/sj.v11i2.2353
Copyright Notice	Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution 4.0 International License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal. 

1. INTRODUCTION

The rapid development of generative artificial intelligence (AI) has transformed academic writing practices in higher education. Among the most widely adopted AI tools, ChatGPT assists users in generating, revising, and refining academic texts with increasing linguistic accuracy and structural coherence. Recent studies suggest that AI-assisted writing can enhance writing efficiency, language quality, and idea organization while simultaneously raising concerns regarding academic integrity, originality, and students' critical engagement with the writing process ([Dwivedi et al., 2023](#); [Kasneci et al., 2023](#)). These opportunities and challenges have stimulated growing interest in understanding how generative AI influences the quality of students' academic writing.

Academic writing is a fundamental competency in higher education because it requires students to construct logical arguments, synthesize evidence, and communicate scholarly ideas using appropriate academic conventions. Developing high-quality academic writing is a complex process that extends beyond linguistic accuracy, requiring critical reading, information evaluation, and coherent argument construction ([Hyland, 2019](#); [Lea & Street, 1998](#)). Recent research has shown that many university students continue to experience difficulties in formulating research problems, organizing ideas, integrating evidence, and maintaining coherence throughout scientific papers, particularly when engaging with discipline-specific academic genres ([Wingate, 2012](#); [Shi et al., 2025](#)). Furthermore, students' academic writing quality is strongly associated with their academic literacy and information literacy, suggesting that effective writing depends not only on language proficiency but also on the ability to locate, evaluate, synthesize, and communicate knowledge critically ([Shi et al., 2025](#); [Lim et al., 2024](#)).

Recent empirical research has increasingly examined the educational use of ChatGPT and other generative AI tools. Existing evidence consistently indicates that AI-assisted writing can improve observable aspects of academic texts, including grammatical accuracy, lexical choice, organizational coherence, and linguistic fluency ([Kasneci et al., 2023](#); [Lo, 2023](#)). At the same time, scholars caution that these benefits are accompanied by important pedagogical and ethical challenges. Effective integration of generative AI requires appropriate lecturer supervision, transparent disclosure of AI assistance, and instructional strategies that encourage students to remain actively engaged in reasoning, source evaluation, and knowledge construction rather than relying excessively on AI-generated content ([Dwivedi et al., 2023](#); [Tlili et al., 2023](#)).

Despite the rapid growth of research on generative AI in education, several important gaps remain. First, existing studies have predominantly examined students' perceptions, acceptance, or general educational impacts of ChatGPT rather than evaluating changes in the quality of scientific writing using structured rubric-based assessments before and after AI-assisted writing ([Lo, 2023](#); [Zhai, 2023](#)). Consequently, empirical evidence linking ChatGPT use to specific dimensions of writing quality remains limited. Second, prior research has generally reported overall improvements in writing performance without distinguishing whether these gains occur primarily in lower-order writing features (e.g., grammar, diction, spelling, and organization) or in higher-order dimensions such as idea development, argumentation, and critical analysis ([Kasneci et al., 2023](#); [Farrokhnia et al., 2024](#)). This distinction is pedagogically important because generative AI may be more effective in supporting linguistic refinement than in fostering deeper conceptual reasoning. Addressing these gaps can provide a more nuanced understanding of which aspects of academic writing benefit most from ChatGPT-assisted revision and can inform evidence-based guidelines for integrating generative AI into higher education writing instruction.

Accordingly, this study addresses the following research question: How does ChatGPT-assisted revision influence the quality of undergraduate students' scientific papers across the dimensions of ideas, organization, grammar, diction, and spelling in a Regional Language and Literature Studies course at an Indonesian language teacher education program in Yogyakarta?

This study contributes to the growing literature on AI-assisted academic writing in three ways. Empirically, it provides rubric-based evidence of changes in students' scientific writing quality before and after ChatGPT-assisted revision. Conceptually, it distinguishes improvements in surface-level writing quality from those related to idea development, thereby offering a more nuanced understanding of AI-supported academic writing. Practically, the findings provide evidence-based recommendations for integrating generative AI into Indonesian language teacher education while maintaining lecturer guidance, academic integrity, and students' critical engagement throughout the writing process.

2. LITERATURE REVIEW

2.1 ChatGPT and Academic Writing Quality

Generative artificial intelligence (AI), particularly ChatGPT, has rapidly emerged as a writing support tool in higher education. Rather than functioning solely as a text generator, ChatGPT assists students throughout multiple stages of the writing process, including planning, drafting, revising, and refining academic texts. Recent empirical studies consistently

demonstrate that AI-assisted writing can improve grammatical accuracy, lexical choice, organizational coherence, and overall writing fluency, enabling students to produce more polished academic texts with greater efficiency ([Dwivedi et al., 2023](#); [Kasneci et al., 2023](#)). Nevertheless, these studies also emphasize that the educational value of ChatGPT depends on responsible pedagogical integration. Excessive reliance on AI may reduce students' active engagement in reasoning, source evaluation, originality, and independent knowledge construction. Consequently, current scholarship increasingly positions ChatGPT as a learning support tool that complements, rather than replaces, students' intellectual work and academic decision-making ([Farrokhnia et al., 2024](#); [Tlili et al., 2023](#)).

In this study, the quality of academic writing is conceptualized as students' ability to produce scientific papers that communicate ideas clearly, logically, and according to academic conventions. Consistent with commonly used analytical writing rubrics, writing quality is assessed through five dimensions: (1) idea development, (2) organization, (3) grammar, (4) diction, and (5) spelling. These dimensions represent both higher-order writing skills, such as developing and organizing ideas, and lower-order language skills, including grammatical accuracy and language mechanics. This distinction provides a more comprehensive evaluation of the aspects of writing that may benefit from ChatGPT-assisted revision.

2.2 Constructivist Perspective on AI-Assisted Writing

This study is informed by constructivist learning theory, which views learning as an active process of knowledge construction through interaction, reflection, and problem solving ([Vygotsky, 1978](#); [Bruner, 1986](#)). Within this perspective, ChatGPT functions as a form of digital scaffolding that supports students during the writing process by providing suggestions for language use, organization, and revision ([Kasneci et al., 2023](#); [Tlili et al., 2023](#)). Nevertheless, constructivism emphasizes that meaningful learning occurs only when students critically evaluate, adapt, and integrate these suggestions into their own reasoning ([Wood et al., 1976](#); [Reiser, 2004](#)). Accordingly, ChatGPT is positioned in this study as a cognitive support tool that facilitates writing development without replacing students' analytical and creative responsibilities ([Dwivedi et al., 2023](#)).

2.3 Academic Literacy and AI-Assisted Writing

Academic literacy encompasses the knowledge and skills required to understand, evaluate, and produce scholarly texts within academic communities ([Lea & Street, 1998](#); [Wingate, 2012](#)). Academic writing quality represents one important manifestation of academic

literacy because it reflects students' ability to communicate ideas using accepted academic conventions ([Hyland, 2019](#)). However, the present study does not directly measure academic literacy as a multidimensional construct ([Lea & Street, 1998](#)). Instead, it evaluates changes in students' academic writing quality through an analytic writing rubric ([Nurgiyantoro, 2016](#)). Therefore, academic literacy serves as the broader conceptual context, whereas writing quality constitutes the empirical focus of the study ([Wingate, 2012](#); [Hyland, 2019](#)).

2.4 Previous Studies and State of the Art

Recent research has consistently demonstrated that ChatGPT and other generative AI tools can support students' academic writing in multiple ways. Across different educational contexts, AI-assisted writing has been associated with improvements in grammatical accuracy, lexical choice, text organization, and overall writing fluency ([Dwivedi et al., 2023](#); [Kasneci et al., 2023](#); [Lo, 2023](#); [Farrokhnia et al., 2024](#)). These findings suggest that generative AI can reduce students' cognitive load during drafting and revision, allowing them to produce more coherent, linguistically accurate, and well-organized academic texts ([Tlili et al., 2023](#); [Farrokhnia et al., 2024](#)).

Despite these benefits, previous studies also emphasize important limitations. While ChatGPT effectively supports surface-level aspects of academic writing, concerns remain regarding students' overreliance on AI, reduced originality, diminished critical engagement, and potential threats to academic integrity ([Dwivedi et al., 2023](#); [Bender et al., 2021](#); [Farrokhnia et al., 2024](#)). Consequently, many scholars argue that ChatGPT should function as a learning support tool under appropriate lecturer supervision rather than as a substitute for students' independent reasoning, knowledge construction, and academic decision-making ([Kasneci et al., 2023](#); [Tlili et al., 2023](#)).

Another stream of research has examined factors influencing students' academic writing beyond AI use. Studies consistently show that the quality of academic writing is shaped by students' academic literacy, information literacy, and disciplinary writing knowledge rather than linguistic proficiency alone ([Lea & Street, 1998](#); [Wingate, 2012](#); [Lim et al., 2024](#)). Research in multilingual contexts further demonstrates that students' first-language (L1) resources influence rhetorical organization, discourse patterns, and linguistic choices in academic writing, particularly when composing in an additional language ([Elosua & Lasagabaster, 2025](#); [Hyland, 2019](#)). Collectively, these studies suggest that academic writing quality reflects the interaction of cognitive, linguistic, and educational factors that extend beyond technological assistance alone.

Although previous studies generally agree that ChatGPT can improve the linguistic quality of students' academic writing ([Kasneci et al., 2023](#); [Lo, 2023](#)), important issues remain unresolved. Existing research has predominantly evaluated students' perceptions of AI, writing productivity, ethical considerations, or overall writing performance ([Dwivedi et al., 2023](#); [Farrokhnia et al., 2024](#)). Comparatively little attention has been given to identifying which specific dimensions of academic writing benefit most from ChatGPT-assisted revision. In particular, few studies distinguish improvements in higher-order writing skills, such as idea development, from improvements in lower-order writing skills, including grammar, diction, spelling, and organization, using an analytic writing rubric ([Kasneci et al., 2023](#); [Tlili et al., 2023](#)). This unresolved issue provides an important direction for further empirical investigation.

2.5 Research Gap and Study Contribution

The existing body of literature provides substantial evidence that ChatGPT can support students' academic writing by improving linguistic accuracy, writing organization, and revision efficiency ([Kasneci et al., 2023](#); [Lo, 2023](#); [Dwivedi et al., 2023](#)). However, three important gaps remain. First, most previous studies have examined students' perceptions of ChatGPT, ethical issues, or overall writing performance rather than evaluating specific dimensions of writing quality through a systematic analytic rubric ([Farrokhnia et al., 2024](#); [Tlili et al., 2023](#)). Second, although improvements in grammar and language mechanics are widely reported, relatively little empirical evidence distinguishes the effects of ChatGPT on higher-order writing skills, such as idea development, from its effects on lower-order writing skills, including organization, diction, grammar, and spelling ([Kasneci et al., 2023](#); [Lo, 2023](#)). Third, evidence from Indonesian language teacher education, particularly within the context of Regional Language and Literature Studies, remains limited.

Addressing these conceptual, methodological, and contextual gaps is important for understanding not only whether ChatGPT improves students' scientific writing, but also which aspects of writing benefit most from AI-assisted revision ([Kasneci et al., 2023](#); [Farrokhnia et al., 2024](#)). Such evidence is needed to inform pedagogical decisions regarding the appropriate integration of generative AI into academic writing instruction while maintaining students' independent reasoning and academic integrity ([Dwivedi et al., 2023](#); [UNESCO, 2023](#)).

This study contributes to the literature in three ways. Empirically, it provides rubric based evidence of changes in students' scientific writing quality before and after ChatGPT assisted revision ([Lim et al., 2024](#); [Zhai, 2023](#)). Methodologically, it employs a before and after analytic

writing assessment that distinguishes idea development from lower-order writing features, including organization, grammar, diction, and spelling, thereby offering a more nuanced evaluation of AI-assisted writing than studies relying primarily on perceptions or overall writing performance ([Crompton and Burke, 2024](#); [Cotton et al., 2024](#)). Practically, the findings provide evidence-based recommendations for integrating generative AI into Indonesian language teacher education while preserving lecturer guidance, students' active engagement in writing, and academic integrity ([UNESCO, 2023](#); [Bearman and Ajjawi, 2023](#)).

3. METHODOLOGY

3.1 Research Design

This study employed a one-group before–after design using document-based assessment and descriptive quantitative analysis ([Creswell & Creswell, 2023](#)). The design compared the quality of scientific papers produced by the same group of students before and after ChatGPT-assisted revision. Because all participants received the same instructional treatment and no control group was included, the findings describe changes in writing quality associated with ChatGPT use rather than establishing a causal relationship between ChatGPT and writing improvement ([Shadish et al., 2002](#)).

The study examined changes in five dimensions of academic writing quality idea development, organization, grammar, diction, and spelling using an analytic writing rubric adapted from [Nurgiyantoro \(2016\)](#). To complement the quantitative findings, qualitative observations of recurring writing patterns, strengths, and weaknesses were conducted during rubric-based assessment. These observations served only as supporting evidence for interpreting changes in writing quality and were not analyzed as an independent qualitative dataset.

3.2 Participants

The study was conducted at “University A,” a private university in Indonesia. This pseudonym protects institutional confidentiality and reduces the risk of indirect identification. Participants were all 30 undergraduate students enrolled in one intact class within a language teacher education program. They attended an undergraduate language and literature course during the 2025/2026 academic year. A total sampling approach was employed because every eligible student receiving the instructional intervention participated in the study ([Creswell & Creswell, 2023](#)). Before the study, participants had completed an introductory academic writing course. They were therefore familiar with basic scientific-writing conventions before completing the two assessed assignments. Since the study involved one class at one institution,

its findings remain context-dependent. They should not be interpreted as representative of broader student populations.

3.3 Study Procedure and Data Collection

The study was conducted from September 2025 to January 2026 using a pretest–posttest design to examine changes in students’ scientific writing performance following ChatGPT-assisted revision. In the first stage (September–October 2025), 30 students produced scientific papers without AI assistance as baseline data. After receiving instruction on the educational use of ChatGPT, students used the tool as a writing-support resource to improve language accuracy, organization, diction, and clarity while maintaining authorship and critically evaluating AI-generated suggestions. This approach follows the principle of human–AI collaboration, in which AI functions as a support tool rather than a replacement for learners’ cognitive engagement ([Kasneci et al., 2023](#); [Tlili et al., 2023](#)).

In the second stage (November–December 2025), students completed a second scientific paper using ChatGPT-assisted revision under lecturer supervision. A total of 60 papers (30 pre-intervention and 30 post-intervention manuscripts) were evaluated using the same analytical writing rubric. The findings were interpreted cautiously because improvements may also have been influenced by students’ increased writing experience, lecturer feedback, and continuous classroom instruction, which are common considerations in educational intervention studies ([Creswell & Creswell, 2023](#))

3.4 Research Instrument

The primary research instrument was an analytic rubric for assessing the quality of students’ scientific papers. The rubric was adapted from [Nurgiyantoro \(2016\)](#) and aligned with the scientific writing guidelines of the institution’s scientific writing guidelines. To ensure consistency, the same rubric was applied to all papers produced before and after ChatGPT-assisted writing.

The rubric evaluates five dimensions of academic writing: (1) idea development, (2) organization, (3) grammar, (4) diction, and (5) spelling. Each dimension contributes to the overall score of 100 points, with the following weight distribution: idea development (35 points), organization (25 points), grammar (20 points), diction (15 points), and spelling (5 points). The higher weighting assigned to idea development and organization reflects their importance in evaluating higher-order academic writing skills, whereas grammar, diction, and spelling represent lower-order language accuracy. Table 1 presents the assessment rubric used in this study.

Table 1. Analytic Rubric for Assessing Scientific Papers

No	Assessment Dimension	Maximum Score	Assessment Focus
1	Idea Development	35	Relevance, originality, clarity, and completeness of ideas
2	Organization	25	Logical structure, coherence, and paragraph organization
3	Grammar	20	Grammatical accuracy and sentence construction
4	Diction	15	Appropriateness and precision of vocabulary
5	Spelling	5	Compliance with Indonesian spelling conventions
Total		100	

Source: Adapted from [Nurgiyantoro \(2016\)](#)

Following assessment, each paper was classified into one of five quality categories adapted from [Nurgiyantoro \(2016\)](#): Excellent (86–100), Good (68–85), Fair (48–67), Poor (32–47), and Very Poor (0–31). These categories were used to describe changes in students’ writing quality before and after ChatGPT-assisted writing.

All scientific papers were evaluated by the researcher using the same analytic rubric. Because the assessment was conducted by a single rater, inter-rater reliability could not be established. This constitutes a methodological limitation and should be considered when interpreting the findings. Nevertheless, the use of a standardized rubric and identical assessment criteria for all papers was intended to maintain scoring consistency throughout the study.

The complete analytic rubric, including assessment descriptors for each writing dimension, is provided in Table 2 to facilitate transparency and replication of the assessment procedure. The analytic rubric used in this study was adapted from [Nurgiyantoro \(2016\)](#) and aligned with the scientific University “A” writing guidelines. The rubric consists of five assessment dimensions with a maximum total score of 100 points.

Table 2. Analytic Rubric for Assessing Scientific Papers

No	Assessment Dimension	Excellent	Good	Fair	Poor
1	Idea Development (35 points)	Ideas are original, relevant, comprehensive, and supported by convincing arguments.	Ideas are relevant and generally well developed but lack depth or completeness.	Ideas are partially developed, with limited supporting arguments.	Ideas are unclear, fragmented, or largely unsupported.
2	Organization (25 points)	Logical organization with clear introduction, body, conclusion, and strong coherence between paragraphs.	Organization is generally logical with minor coherence problems.	Organization is inconsistent, and transitions between paragraphs are weak.	Organization is unclear, making the paper difficult to follow.
3	Grammar (20 points)	Grammar is consistently accurate with only negligible errors that do not affect comprehension.	Minor grammatical errors occur but do not interfere with understanding.	Frequent grammatical errors occasionally affect clarity.	Persistent grammatical errors significantly interfere with communication.
4	Diction (15 points)	Vocabulary is accurate, varied, and appropriate for academic writing.	Vocabulary is generally appropriate with occasional inaccuracies.	Vocabulary is repetitive or sometimes inappropriate.	Vocabulary is limited and frequently

No	Assessment Dimension	Excellent	Good	Fair	Poor
5	Spelling (5 points)	Correct spelling and punctuation throughout the paper.	Very few spelling or punctuation errors.	Several spelling or punctuation errors.	inappropriate for academic writing. Frequent spelling and punctuation errors reduce readability.

Source: Adapted from [Nurgiyantoro \(2016\)](#)

3.5 Scoring Procedure

Each scientific paper was assessed analytically by evaluating all five dimensions separately. The scores from the five dimensions were summed to obtain a maximum score of 100 points. The same rubric was consistently applied to all papers produced before and after ChatGPT-assisted writing.

3.6 Quality Classification

The classification of scientific paper quality was established to provide a systematic framework for interpreting students’ writing performance. The categories were determined based on the total scores obtained from the assessment rubric, representing different levels of writing quality. Table 3 presents the criteria for scientific paper quality categories used in evaluating the participating students writing performance.

Table 3. Scientific Paper Quality Categories

No	Score Range	Category
1	86–100	Excellent
2	68–85	Good
3	48–67	Fair
4	32–47	Poor
5	0–31	Very Poor

Source: Adapted from [Nurgiyantoro \(2016\)](#).

3.7 Data Analysis

The data was analyzed using descriptive quantitative analysis supported by qualitative interpretation. Quantitative analysis focused on comparing the mean scores of students’ scientific papers before and after ChatGPT-assisted writing. Mean scores were calculated for the overall writing quality and for each of the five assessment dimensions: idea development, organization, grammar, diction, and spelling. Descriptive statistics, including mean scores and frequency distributions, are commonly used in educational research to summarize patterns of learner performance and describe changes across measurement points ([Creswell & Creswell, 2023](#); [Ary et al., 2021](#)).

The distribution of papers across the five quality categories (Excellent, Good, Fair, Poor, and Very Poor) was also calculated to describe changes in students’ writing performance after the use of ChatGPT. Because the study employed a one-group before–after design without a control group, the analysis was limited to descriptive comparisons rather than inferential

statistical testing. Such a design is appropriate for identifying patterns of change within a single group; however, the absence of a comparison group limits conclusions regarding causal effects ([Creswell & Creswell, 2023](#); [Shadish et al., 2002](#)). Therefore, the results indicate changes associated with ChatGPT-assisted writing and should not be interpreted as evidence of a causal relationship.

To complement the quantitative findings, qualitative observations were conducted by repeatedly examining students' papers to identify recurring patterns of improvement and persistent weaknesses across the five assessment dimensions. This iterative document analysis followed the interactive qualitative data analysis approach proposed by [Miles et al. \(2014\)](#), which involves data condensation, data display, and conclusion drawing. Document analysis is widely recognized as a systematic method for examining written artifacts to identify patterns, meanings, and contextual information ([Bowen, 2009](#)).

Finally, quantitative and qualitative findings were interpreted together to provide a comprehensive description of changes in students' scientific writing quality before and after ChatGPT-assisted writing. The integration of quantitative trends and qualitative interpretation enables researchers to develop a more complete understanding of educational phenomena by combining numerical evidence with contextual explanations ([Creswell & Plano Clark, 2018](#)).

4. FINDINGS

4.1 Research Results

This study compared the quality of scientific papers written by 30 student papers before and after ChatGPT assisted writing in the Regional Language and Literature Studies course. The results are presented descriptively using overall scores, quality categories, aspect-level comparisons, and representative writing examples.

4.2 Overall Writing Quality Before and After ChatGPT Use

This section presents the comparative analysis of students' scientific writing quality before and after the integration of ChatGPT as a writing-support tool. The analysis aims to identify changes in students' overall writing performance based on the established assessment criteria for scientific papers. The comparison focuses on the differences in writing quality across five dimensions: idea development, organization, grammar, diction, and spelling. The assessment results provide an overview of how ChatGPT-assisted writing influenced the quality of students' academic texts within the context of Indonesian language education. Table

4 presents the criteria used to evaluate the quality of scientific papers produced by participating students before using ChatGPT.

Table 4 Criteria for Assessing the Quality of Scientific Papers Participating Students Before Using ChatGPT

No. Sample	Criteria for Assessing the Quality of Scientific Papers					Quantity (100)
	1	2	3	4	5	
	0-35	0-25	0-20	0-15	0-5	
1	28	22	15	12	3	80
2	25	20	15	10	2	72
3	25	21	17	12	2	77
4	28	20	17	11	3	79
5	32	21	14	10	2	79
6	33	19	14	10	3	79
7	29	22	16	12	3	82
8	25	23	15	12	3	78
9	27	20	15	11	3	76
10	28	18	13	13	3	75
11	25	21	17	12	2	77
12	28	20	17	11	3	79
13	32	21	14	10	2	79
14	33	19	14	10	3	79
15	29	22	16	12	3	82
16	25	23	15	12	3	78
17	27	20	15	11	3	76
18	28	18	13	13	3	75
19	30	21	12	10	3	76
20	29	22	13	11	3	78
21	31	20	15	11	3	80
22	32	18	17	10	3	80
23	30	19	11	10	3	73
24	33	21	12	12	3	81
25	31	20	12	12	3	78
26	29	20	14	10	3	76
27	25	21	15	11	2	74
28	25	21	17	12	2	77
29	32	21	15	10	4	82
30	30	20	15	12	3	80
Value Average	28,8	20,4	14,6	11,1	2,8	77,9

Source: Research Data (2025)

Table Description:

- 1. Idea
- 2. Organization
- 3. Grammar
- 4. Diction
- 5. Spelling

The percentage distribution of students' scientific paper assessment results before using ChatGPT provides an overview of their initial writing performance. The analysis illustrates the proportion of papers classified into each quality category based on the predetermined assessment criteria. Table 5 presents the percentage of writing quality levels among participating students' scientific papers before the implementation of ChatGPT-assisted writing.

Table 5 Percentage of Assessment for Writing Scientific Papers for Students Before Using ChatGPT

Category	Value	Quantity	Percentage
Excellent	86-100	0	0
Good	68-85	30	100
Enough	48-67	0	0
Less	32-47	0	0
Very less	0-31	0	0

Source: Research Data (2025)

Before using ChatGPT, the overall mean score of students' scientific papers was 77.9, indicating that all papers were classified in the good category (68–85). None of the papers reached the Excellent category, while all 30 papers (100%) were categorized as Good. Individual scores ranged from 72 to 82.

Following the implementation of ChatGPT-assisted writing, students' scientific papers were reassessed using the same evaluation criteria to identify changes in writing quality. The assessment focused on the five dimensions of scientific writing, including idea development, organization, grammar, diction, and spelling. Applying consistent criteria before and after ChatGPT use enabled a direct comparison of students' writing performance and the extent of improvement across different aspects of academic writing. Table 6 presents the criteria used to assess the quality of scientific papers produced by students after using ChatGPT.

Table 6 Criteria for Assessing the Quality of Scientific Students paper After Using ChatGPT

No Sample	Criteria for Assessing the Quality of Scientific Papers					Quantity (100)
	1	2	3	4	5	
	0-35	0-25	0-20	0-15	0-5	
1	32	23	17	14	4	90
2	32	23	17	12	4	88
3	31	23	19	14	4	91
4	34	22	19	13	3	93
5	33	22	17	12	3	87
6	33	22	18	12	4	89
7	32	23	16	14	4	89
8	33	23	18	14	4	92
9	32	21	18	12	4	87

The Quality of Scientific Papers of Undergraduate Students in Regional Language and Literature Studies Lectures Before and After Using ChatGPT as a Writing Tool

Ari Wulandari, Arif Budiman, Tri Ratna Herawati, Primasari Wahyuni, Diaz Danu Setyawan

10	33	21	17	14	4	89
11	30	22	18	14	4	88
12	32	22	10	14	4	82
13	34	23	19	14	4	94
14	34	22	17	13	4	90
15	32	23	19	14	3	91
16	33	24	18	14	3	90
17	32	23	17	14	4	90
18	31	22	18	14	4	89
19	30	22	15	12	4	83
20	32	23	17	12	3	87
21	32	21	17	12	3	85
22	34	23	18	13	3	91
23	32	22	15	14	3	86
24	33	22	17	13	3	88
25	33	21	19	14	4	91
26	32	22	18	14	4	90
27	32	23	18	14	4	91
28	33	23	17	14	4	91
29	33	22	18	12	4	89
30	32	22	18	14	4	90
Value Average	32,3	22,3	17,8	14,2	3,7	89

Source: Research Data (2025)

Table Description:

1. Idea
2. Organization
3. Grammar
4. Diction
5. Spelling

The evaluation results after the implementation of ChatGPT-assisted writing demonstrate the distribution of students' scientific papers across different quality categories. The percentage analysis was conducted to identify shifts in students' writing performance and to illustrate the extent of improvement after using ChatGPT as a writing-support tool. Table 7 presents the percentage distribution of scientific paper quality categories among students after using ChatGPT.

Table 7 Percentage of Scientific Paper Quality Category for Students After Using ChatGPT

Category	Value	Quantity	Percentage
Excellent	86-100	27	90
Good	68-85	3	10
Enough	48-67	0	0
Less	32-47	0	0
Very less	0-31	0	0

Source: Research Data (2025)

After ChatGPT-assisted writing, the overall mean score increased to 89.0, placing the average performance in the Excellent category (86–100). Twenty-seven papers (90%) were classified as Excellent, while the remaining three papers (10%) remained in the Good category. Individual scores ranged from 82 to 94.

4.3 Comparison of Writing Quality Before and After ChatGPT Use

The comparison of mean scores across writing dimensions provides a more detailed view of changes in students' scientific writing performance before and after ChatGPT-assisted writing. The analysis examines the extent of improvement in each dimension, including idea development, organization, grammar, diction, and spelling. Comparing the mean scores allows the researchers to identify which aspects of writing experienced the most significant improvement following the integration of ChatGPT as a writing-support tool. Table 8 presents the comparison of students' mean scores across the five assessed writing dimensions before and after using ChatGPT.

Table 8. Comparison of Mean Scores Across Writing Dimensions

No.	Assessment Dimension	Before	After	Gain	Percentage Gain
1	Idea Development	28.8	32.3	+3.5	12.2%
2	Organization	20.4	22.3	+1.9	9.3%
3	Grammar	14.6	17.8	+3.2	21.9%
4	Diction	11.1	14.2	+3.1	27.9%
5	Spelling	2.8	3.7	+0.9	32.1%
Overall Score	77.9	89.0	+11.1	14.2%	

Source: Research Data (2025)

Table 8 shows improvements across all five assessment dimensions after ChatGPT-assisted writing. The largest relative improvements were observed in spelling (32.1%), diction (27.9%), and grammar (21.9%). Improvements in idea development (12.2%) and organization (9.3%) were comparatively smaller. Overall, the mean writing score increased by 11.1 points, representing a 14.2% improvement over the initial average score.

4.4 Changes in Quality Categories

The changes in scientific paper quality categories illustrate the overall shift in students' writing performance after the implementation of ChatGPT assisted writing. This analysis focuses on the movement of students' papers across the five quality levels, namely Excellent, Good, Fair, Poor, and Very Poor, based on the established assessment criteria. The comparison of category distributions provides an overview of the extent to which ChatGPT assisted writing contributed to improvements in students' scientific writing quality. Table 9 presents the changes in scientific paper quality categories before and after the use of ChatGPT by University "A" students.

Table 9. Changes in Scientific Paper Quality Categories

No.	Quality Category	Before	After	Change
1	Excellent	0 (0%)	27 (90%)	+27
2	Good	30 (100%)	3 (10%)	-27
3	Fair	0	0	0
4	Poor	0	0	0
5	Very Poor	0	0	0

Source: Research Data (2025)

As shown in Table 9, the quality distribution shifted substantially following ChatGPT-assisted writing. Before using ChatGPT, all students were classified in the good category. After using ChatGPT, 27 students (90%) moved into the Excellent category, while only three students (10%) remained in the good category. No papers were classified as Fair, Poor, or Very Poor in either assessment.

4.5 Representative Example of Writing Improvement

This section presents representative examples of students' writing to illustrate qualitative changes that occurred after the implementation of ChatGPT assisted writing. The examples were selected to demonstrate observable improvements in several writing dimensions, including idea development, organization, language accuracy, diction, and spelling. By comparing excerpts from students' papers before and after ChatGPT use, this analysis provides contextual evidence supporting the quantitative findings presented in the previous sections. Table 10 presents representative examples of writing improvements in students' scientific papers before and after ChatGPT-assisted writing.

Table 10. Representative Examples of Writing Improvements Before and After ChatGPT-Assisted Writing

No	Writing Aspect	Before ChatGPT	After ChatGPT	Improvement Observed
1	Idea Development	<i>Regional languages are important because they are still used today.</i>	<i>Regional languages play an essential role in preserving cultural identity because they function as a medium for transmitting local knowledge, traditions, and values across generations.</i>	Ideas became more elaborated, specific, and academically developed.
2	Organization	<i>This paper discusses regional languages. There are many theories. The method is qualitative. The results are explained afterwards.</i>	<i>This paper begins by introducing the importance of regional language preservation, followed by a review of relevant theories, a description of the qualitative methodology, presentation of findings, and concluding remarks.</i>	Paragraph organization and logical sequencing became clearer and more coherent.
3	Grammar	<i>Regional language have many function in society and still used by people.</i>	<i>Regional languages have many functions in society and are still widely used by local communities.</i>	Improved subject–verb agreement, sentence structure, and grammatical accuracy.

4	Diction	<i>The data is very good and shows many interesting things.</i>	<i>The findings indicate several significant patterns regarding the preservation and use of regional languages within the studied community.</i>	Informal vocabulary was replaced with more precise academic expressions.
5	Spelling	<i>Based on the research, the students can write better. The result also show improvement</i>	<i>Based on the research, the students were able to produce higher-quality scientific papers. The results also showed measurable improvement.</i>	Correct punctuation, capitalization, spacing, and verb form enhanced readability.

Source: Representative excerpts adapted from students' scientific papers (2025).

The representative excerpt illustrates improvements in grammatical accuracy, academic vocabulary, sentence structure, and overall readability following ChatGPT-assisted writing. Similar improvements were consistently observed across many of the assessed papers.

Overall, the findings demonstrate consistent improvements in students' scientific writing after ChatGPT-assisted writing. Improvements were observed across all five assessment dimensions, with the greatest gains occurring in spelling, diction, and grammar. In addition, the distribution of writing quality shifted substantially, as 90% of students moved from the Good category to the Excellent category. The interpretation of these findings in relation to constructivism, academic literacy, digital scaffolding, and previous studies is presented in the Discussion section.

5. DISCUSSION

5.1 Interpretation of the Main Findings

The present study found that the quality of students' scientific papers improved across all five dimensions of the assessment rubric after ChatGPT-assisted writing, namely idea development, organization, grammar, diction, and spelling. The overall mean score increased from 77.9 to 89.0, and 27 of the 30 papers (90%) shifted from the Good to the Excellent category. These findings suggest that ChatGPT may function as a supportive tool in academic writing processes by assisting students in generating ideas, improving textual organization, and refining language accuracy through revision and feedback mechanisms ([Kasneci et al., 2023](#); [Tlili et al., 2023](#)).

The improvement observed across the assessed dimensions is consistent with previous research indicating that automated writing support tools can enhance students' revision practices and contribute to improvements in linguistic accuracy and text quality when used critically and interactively ([Ranalli, 2021](#); [Zhang & Hyland, 2023](#)). However, because these

five dimensions constituted the assessment instrument, the discussion is limited to interpreting improvements in observable writing quality rather than broader educational constructs that were not directly measured. This limitation follows methodological principles in educational research that emphasize the importance of aligning interpretations with the scope and validity of the measurement instrument used ([Creswell & Creswell, 2023](#); [Messick, 1995](#)).

5.2 Improvement in Technical Writing Quality

The largest improvements were observed in grammar, diction, and spelling. These findings suggest that ChatGPT is particularly effective in supporting technical aspects of academic writing, including sentence accuracy, lexical appropriateness, and conformity to academic language conventions. Similar observations have been reported by [Kasneci et al. \(2023\)](#) and [Dwivedi et al. \(2023\)](#), who found that large language models facilitate language editing, provide linguistic feedback, and improve textual fluency.

However, the present study extends previous research by demonstrating that these improvements were measured through an analytic writing rubric applied to authentic student papers before and after ChatGPT-assisted writing rather than relying on students' perceptions or self-reported experiences. This provides empirical evidence that improvements are observable in specific writing dimensions rather than merely perceived by users ([Creswell & Creswell, 2023](#); [Messick, 1995](#)).

5.3 Unequal Improvements Across Writing Dimensions

Although all assessment dimensions improved, gains were not evenly distributed. Improvements in idea development and organization were smaller than those observed for grammar, diction, and spelling. This pattern suggests that ChatGPT contributes more strongly to linguistic refinement than to conceptual development, as large language models are generally more effective in supporting language generation, revision, and surface-level textual improvement than complex reasoning and knowledge construction ([Kasneci et al., 2023](#); [Dwivedi et al., 2023](#)).

Importantly, the present study did not directly assess critical thinking, originality, or academic literacy. Therefore, no conclusion can be drawn regarding students' critical thinking abilities or the originality of their work. Instead, the findings simply indicate that improvements in the rubric dimensions related to ideas and organization were more modest than improvements in language-related dimensions. Future research employing dedicated measures of higher-order thinking is needed to examine these broader educational outcomes ([Creswell & Creswell, 2023](#); [Facione, 2020](#)).

5.4 Comparison with Previous Studies

The present findings generally support earlier studies reporting that ChatGPT improves academic writing quality ([Lo, 2023](#); [Kasneci et al., 2023](#)). However, this study also provides a more nuanced perspective. Previous studies often describe ChatGPT as improving academic writing in general, whereas the present findings demonstrate that improvements vary considerably across individual writing dimensions. Technical language features showed substantially larger gains than conceptual aspects of writing. This suggests that writing quality should not be viewed as a single construct, but rather as multiple dimensions that may respond differently to AI-assisted writing ([Hyland, 2019](#); [Barkaoui, 2023](#)).

These findings also qualify concerns raised by [Bender et al. \(2021\)](#). While their work cautions that large language models may generate fluent texts without genuine conceptual understanding, the present study neither confirms nor rejects this concern because conceptual understanding was not directly measured. Instead, the findings indicate that observable improvements occurred primarily in language quality, while improvements in idea development remained comparatively modest. This interpretation aligns with recent discussions suggesting that generative AI tools can enhance textual production while requiring human judgment to ensure originality, meaning construction, and deeper academic engagement ([Kasneci et al., 2023](#); [Dwivedi et al., 2023](#)).

5.5 Pedagogical Implications

The unequal improvement across writing dimensions has important implications for AI-assisted writing instruction. If ChatGPT primarily supports language accuracy and textual organization, lecturers may need to redesign writing assignments to emphasize reasoning, evidence evaluation, source integration, and contextual interpretation rather than sentence production alone ([Kasneci et al., 2023](#)). In courses such as Regional Language and Literature Studies, assignments may increasingly require students to interpret local cultural phenomena, compare theoretical perspectives, and justify arguments using evidence that extends beyond AI-generated text. Such instructional strategies may help maximize the benefits of ChatGPT while maintaining students' independent intellectual engagement and academic integrity ([Dwivedi et al., 2023](#); [UNESCO, 2023](#)).

5.6 Strengths and Limitations

This study has several strengths. First, it employed a one-group before–after document comparison that enabled direct observation of changes in students' writing products using the same analytic rubric. Second, all papers were evaluated using identical assessment criteria

covering five dimensions of scientific writing, allowing systematic comparison across writing aspects. Third, the study contributes empirical evidence from an Indonesian higher education context that has received limited attention in previous ChatGPT research ([Kasneci et al., 2023](#); [Dwivedi et al., 2023](#)).

Several limitations should also be acknowledged. First, the study used a one-group before–after design without a control group. Consequently, improvements cannot be attributed exclusively to ChatGPT because students also received regular classroom instruction, lecturer feedback, and additional writing experience during the semester ([Creswell & Creswell, 2023](#)). Second, the scientific papers were assessed by a single rater using an analytic rubric; therefore, inter-rater reliability was not established. Third, the assessment focused exclusively on observable writing quality (idea development, organization, grammar, diction, and spelling). Broader constructs such as critical thinking, academic literacy, originality, and long-term writing independence were beyond the scope of the instrument and were therefore not evaluated ([Messick, 1995](#)). Future studies employing experimental designs, multiple raters, and additional measures of higher-order writing abilities would provide stronger evidence regarding the educational impact of ChatGPT-assisted writing.

5.7 Contribution of the Study

Rather than demonstrating that ChatGPT universally improves academic writing, this study contributes a more differentiated understanding of AI-assisted writing by showing that improvements are distributed unevenly across specific dimensions of writing quality. The findings indicate that language-related aspects benefit more substantially than idea development and organization. This distinction refines current understanding of ChatGPT-assisted academic writing and suggests that future evaluations should examine individual writing dimensions rather than relying solely on overall writing scores ([Kasneci et al., 2023](#); [Dwivedi et al., 2023](#)). Such dimension-specific analysis may provide a more accurate account of how generative AI influences academic writing processes and outcomes ([Zhang & Hyland, 2023](#)).

6. CONCLUSION

This study found that ChatGPT assisted writing was associated with improvements in the quality of students' scientific papers, as reflected by an increase in the overall mean score from 77.9 to 89.0 and improvements across the five assessed writing dimensions: idea development, organization, grammar, diction, and spelling. The largest gains were observed in grammar, diction, and spelling, whereas improvements in idea development and organization were

comparatively smaller. These findings suggest that, within the context of this study, ChatGPT was particularly useful for supporting the linguistic and structural aspects of academic writing rather than uniformly improving all dimensions of writing quality.

The unequal pattern of improvement indicates that ChatGPT may function most effectively as a writing-support tool for language refinement and text organization. Because this study assessed only observable writing quality, no conclusions can be drawn regarding broader educational constructs such as critical thinking, originality, or academic literacy. Accordingly, ChatGPT should be viewed as a complementary learning resource that supports students' writing process while leaving the development of ideas, argumentation, and independent reasoning to students themselves.

This study has several limitations that should be considered when interpreting the findings. The research employed a one-group before–after design without a control group; therefore, the observed improvements cannot be attributed exclusively to ChatGPT-assisted writing. In addition, the study involved a single class within one study program, the papers were evaluated by a single rater without inter-rater reliability testing, and the analysis relied primarily on descriptive comparisons. Consequently, the findings should be interpreted as evidence of changes in writing quality associated with ChatGPT use rather than as definitive proof of causal effectiveness.

Future research should employ controlled experimental designs, larger and more diverse samples, multiple raters, and inferential statistical analyses to strengthen the evidence on AI-assisted academic writing. From a practical perspective, lecturers and higher education institutions are encouraged to establish clear guidelines for responsible AI use by requiring students to disclose AI assistance, limiting ChatGPT primarily to drafting revision and language editing, and incorporating rubric-guided reflection activities that encourage students to evaluate, revise, and justify AI-generated suggestions independently. Such practices may help maximize the educational benefits of ChatGPT while maintaining academic integrity and students' independent intellectual development.

DECLARATION OF GENERATIVE AI

The authors used AI-assisted tools, including ChatGPT, Grammarly, and QuillBot, to support language editing, grammar checking, and clarity improvement during manuscript preparation. The authors carefully reviewed, revised, and approved all AI-assisted outputs and take full responsibility for the content, accuracy, and integrity of the manuscript.

7. REFERENCES

- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2021). *Introduction to research in education* (10th ed.). Cengage Learning.
- Barkaoui, K. (2023). Examining the impact of automated writing evaluation on second language writing: A systematic review. *Assessing Writing*, 55, 100685. <https://doi.org/10.1016/j.asw.2022.100685>
- Bearman, M., & Ajjawi, R. (2023). AI and the future of assessment in higher education. *Medical Education*, 57(11), 987–989. <https://doi.org/10.1111/medu.15145>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Bruner, J. (1986). *Actual minds, possible worlds*. Harvard University Press.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Crompton, H., & Burke, D. (2024). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 21, 22. <https://doi.org/10.1186/s41239-024-00439-9>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M., Al-Busaidi, K. A., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elosua, P., & Lasagabaster, D. (2025). The role of first language resources in multilingual academic writing. *Journal of Second Language Writing*, 67, 101170. <https://doi.org/10.1016/j.jslw.2024.101170>
- Facione, P. A. (2020). *Critical thinking: What it is and why it counts* (7th ed.). Insight Assessment.

- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. E. J. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Hyland, K. (2019). *Second language writing* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108509383>
- 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., Pfeiffer, F., 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, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157–172. <https://doi.org/10.1080/03075079812331380364>
- Lim, W. M., Gunasekara, A. N., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2024). Generative AI and the future of education: A systematic review of empirical evidence. *International Journal of Educational Technology in Higher Education*, 21, Article 20. <https://doi.org/10.1186/s41239-024-00455-9>
- Lo, C. K. (2023). *What is the impact of ChatGPT on education? A rapid review of the literature*. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Nurgiyantoro, B. (2016). *Penilaian pembelajaran bahasa berbasis kompetensi* (2nd ed.). BPFE.
- Ranalli, J. (2021). L2 student engagement with automated written corrective feedback: Current status and future directions. *Language Teaching Research*, 25(5), 730–749. <https://doi.org/10.1177/1362168820934230>
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

- Shi, J., Zhang, Y., Liu, X., & Wang, H. (2025). Academic literacy, information literacy, and university students' academic writing quality. *Journal of Writing Research*, 17(1), 45–67.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). *What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education*. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11(2), 145–154. <https://doi.org/10.1016/j.jeap.2011.11.001>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zhai, X. (2023). ChatGPT for next generation science learning. *Journal of Research in Science Teaching*, 60(8), 1524–1530. <https://doi.org/10.1002/tea.21816>
- Zhang, Z. V., & Hyland, K. (2023). Student engagement with automated feedback on academic writing: Revising, responding, and learning. *Assessing Writing*, 55, 100694. <https://doi.org/10.1016/j.asw.2022.100694>