Assessing Graduate Students' Perceptions in Exploring Generative AI for Scientific Writing

Paulo César Polastri[1,3], **Flávia Linhalis**[1], **Julio Cesar dos Reis**[1,2]

[1]Núcleo de Informática Aplicada à Educação (NIED), UNICAMP, Brazil

[2]Instituto de Computação (IC), UNICAMP, Brazil

[3]Instituto Federal de Educação, Ciência e Tecnologia (IFSP), Brazil

paulo.polastri@ifsp.edu.br,{flalin,dosreis}@unicamp.br

Abstract. *Graduate students often face challenges in scientific writing. While Generative AI (GenAI) offers support, there is little empirical evidence on students' perceptions of how they use GenAI tools for this purpose and the way tutorials can help. This study investigates graduate students' perceptions of the use of GenAI to support scientific writing. We conducted an assessment with Graduate students in Computer Science using GenAI tools via provided tutorials. The results indicate positive evaluations of the tutorials' clarity and overall quality. Students clearly view GenAI as an assistant rather than as a substitute for human authorship. Our findings suggest that the designed tutorials can complement instruction in scientific writing in graduate school.*

1. Introduction

Scientific writing is a fundamental component of graduate education and the primary means by which researchers communicate their findings and advance knowledge. Graduate students frequently encounter challenges in developing scientific documents, including structuring arguments, ensuring methodological rigor, mastering academic conventions, and articulating complex ideas clearly [Inouye and McAlpine 2019, Lonka et al. 2019]. These difficulties can negatively affect academic outcomes, delay degree completion, and reduce confidence in pursuing research careers [Maher et al. 2014, Odena and Burgess 2017].

In recent years, artificial intelligence, particularly Generative AI (GenAI), has transformed the educational landscape. Large Language Models (LLMs) have been increasingly adopted in academic contexts, providing writing assistance, explanations, and quick access to information [Kasneci et al. 2023, Su and Yang 2023]. These technologies offer opportunities for scalable and personalized support, complementing traditional educational resources and addressing individual learning needs [Peláez-Sánchez et al. 2024, Dwivedi et al. 2023, Sullivan et al. 2023]. Prior studies indicate that students perceive GenAI as beneficial for multiple aspects of their learning [Daza et al. 2024].

Despite the availability of writing guides and workshops, graduate students continue to face significant obstacles in mastering scientific writing [Scalfani et al. 2016, Fanguy et al. 2019]. Common difficulties include structuring introductions, identifying research gaps, justifying methodological decisions, and synthesizing results into coherent discussions [Pololi et al. 2011]. These challenges are intensified by disciplinary differences in writing conventions and limited access to personalized feedback, particularly in programs with high student-to-faculty ratios [North 2005, Lax 2014, McAlpine and Amundsen 2011, Gopee and Deane 2013]. Without adequate guidance, students may struggle to develop writing proficiency effectively [Lu and Wang 2022, McLaughlin et al. 2013].

GenAI offers potential solutions to these challenges through GenAI-guided tutorials on scientific writing content, tailored to specific sections of scientific documents. Such tutorials can provide structural guidance, examples, and explanations adapted to different levels of understanding, supporting the development of scientific writing skills [Su and Yang 2023, Daza et al. 2024, Bhutoria 2022, Hwang 2014]. We observe that the literature remains limited on how graduate students perceive and use GenAI tools in their tasks, and on the benefits and limitations of designed tutorials in practice [Tlili et al. 2023, van Dis et al. 2023]. Understanding these experiences is essential for developing effective pedagogical strategies that integrate GenAI while maintaining academic rigor [Kasneci et al. 2023].

This study investigates the development and evaluation of AI-generated tutorials for supporting scientific writing among graduate students. Our objective is to create targeted tutorials for key sections of academic documents and to assess students' perceptions of their usefulness, quality, and impact on the writing process and on their learning outcomes. To this end, we conducted an evaluation with graduate students in a computer science course in Brazil.

This research offers three main contributions. First, it proposes developing AI-guided tutorials to support scientific writing. Second, it provides empirical evidence on graduate students' perceptions and how these materials influence their understanding, confidence, and learning experience. Third, it identifies the benefits and challenges of generative AI in creating educational and scientific content, offering insights into future AI-assisted learning tools and curriculum integration [Kasneci et al. 2023, Tlili et al. 2023].

This article is organized as follows: Section 2 presents the theoretical background, covering fundamental concepts related to scientific writing in graduate education and reviews related work; Section 3 describes our study design; Section 4 presents the results, analyzing students' perceptions. Section 5 discusses our findings and examines the study's limitations. Finally, Section 6 presents concluding remarks.

2. Background and Related Work

Scientific writing is essential in graduate education, enabling students to communicate research, contribute to knowledge, and develop academic identities. However, formal training is often limited [Gupta et al. 2022], leaving students to develop skills through practice over their careers. This process shapes how they express arguments, methodologies, and theoretical perspectives.

Graduate students commonly face difficulties in managing the writing process, communicating ideas, and understanding academic expectations [McRell et al. 2021]. Scientific writing is particularly demanding due to its complex structure [Lin and Morrison 2021]. Reported challenges include limited knowledge of writing conventions, low proficiency in academic language, reduced motivation, and difficulties with foreign-language competence [Cahill et al. 2008]. These challenges also burden supervisors and contribute to inconsistent performance, reinforcing the need for stronger pedagogical support [Gupta et al. 2022]. Consequently, many students rely on inefficient approaches that may reinforce misconceptions [Shah et al. 2009].

Scientific documents follow established conventions, such as the IMRAD structure (Introduction, Methods, Results, Discussion), which is widely used in research articles [Kasneci et al. 2023]. Although useful, this structure can be overly rigid, requiring a deeper understanding to be applied effectively. The literature highlights recurring challenges, including reliance on model papers [Shah et al. 2009] and high cognitive burden during writing tasks [Peña-Lima et al. 2026]. Students must simultaneously manage content knowledge, rhetorical organization, and language accuracy.

Selecting and citing relevant literature is also difficult, as it requires experience and the ability to identify original contributions amid extensive research [Lin and Morrison 2021]. These issues demonstrate that scientific writing involves not only linguistic skills but also disciplinary knowledge and strategic thinking.

GenAI has transformed natural language processing. LLMs, based on deep learning and trained on large datasets, can understand and generate human language across tasks [Vaswani et al. 2017, Brown et al. 2020]. Their development relies on transformer architectures, which use self-attention to process sequences in parallel and capture complex patterns [Vaswani et al. 2017, Luitse and Denkena 2021].

This innovation enabled models to capture long-range relationships in text [Luitse and Denkena 2021]. Models such as BERT and GPT, LLMs, have advanced rapidly [Mienye et al. 2025]. For instance, since the release of ChatGPT, GenAI has been increasingly used in education to support writing, enhance understanding, and assist with problem-solving [Lo 2023, Tlili et al. 2023]. Potential benefits include more accessible, personalized, and interactive learning experiences [Kasneci et al. 2023]. However, concerns remain. Over-reliance may reduce critical thinking, originality, and independent learning, while promoting uncritical acceptance of AI outputs [Kasneci et al. 2023, Tlili et al. 2023, Dwivedi et al. 2023]. Additional issues include data privacy and algorithmic bias [Sullivan et al. 2023, van Dis et al. 2023].

GenAI can support scientific writing by assisting with organization, drafting, proofreading, and idea generation, allowing users to focus on higher-level tasks [Lendvai 2025, Imran and Almusharraf 2023, Tlili et al. 2023]. Compared to earlier systems, it offers better contextual understanding, coherence, and personalized feedback [Song and Song 2023]. Nevertheless, optimal results occur when AI support is combined with instructor guidance, reinforcing its role as a complement rather than a replacement for human teaching [Kasneci et al. 2023].

Research has examined the effectiveness of GenAI tools in scientific writing workflows. A narrative review by [Granjeiro et al. 2025] found that platforms such as Elicit and Consensus are particularly effective for literature reviews and identifying research gaps, while ChatGPT supports structural drafting and summarization, and Grammarly improves stylistic consistency and plagiarism detection. However, limitations remain, including the risk of fabricated citations (“hallucinations”) and insufficient contextual understanding of specialized terminology, reinforcing the need for human oversight to ensure technical accuracy and bibliographic integrity.

Similarly, [Ren and Wang 2025] proposed a quantitative framework to evaluate LLM-generated research proposals, showing that iterative prompting improves structural coherence and academic alignment. Despite these advances, a persistent "reference validity gap" remains, as models occasionally generate inaccurate or nonexistent citations, and current methods still struggle to assess the intellectual novelty of AI-generated research questions. In a comparative study, [Shopovski et al. 2025] found that the ChatGPT tool provided useful feedback on language clarity and structure when reviewing manuscripts, yet demonstrated lower critical rigor than human reviewers, often overlooking methodological flaws and favoring minor revisions. This identified a "rigor gap", indicating limited capability for deep scientific evaluation.

Furthermore, [Kosmyna et al. 2025] investigated AI use in complex doctoral writing tasks and found that, although such tools reduce writer's block and enhance lexical richness, they may also decrease critical engagement with content. This phenomenon, described as "cognitive debt", suggests a trade-off between writing efficiency and depth of argumentation. The study also highlighted concerns about potential deskilling among early-career researchers and emphasized the lack of evidence on the long-term cognitive effects of reliance on AI.

Overall, while GenAI has demonstrated considerable potential to support the construction of scientific texts, important challenges remain regarding methodological rigor, intellectual originality, and the preservation of independent critical thinking. Rather than positioning GenAI as a replacement for scholarly reasoning, our present study frames it as a structured pedagogical aid designed to scaffold the development of scientific writing skills. By explicitly addressing the conceptual and methodological dimensions of research communication, our approach aims to guide graduate students in understanding how scientific arguments are constructed, how research gaps are articulated, and how evidence is synthesized. In this sense, this study contributes to leveraging GenAI for scientific text writing and to understanding how AI-assisted guidance can foster deeper learning processes while maintaining the epistemic standards expected in academic research. In addition, we provide original quantitative and qualitative analyses from students' generated data in a real-world graduate course in Computer Science.

3. Study Design

The study was conducted during a graduate-level Scientific Writing course offered at the Institute of Computing, University of Campinas (UNICAMP)[1], Brazil, with nine students ($n = 9$), and evaluated by participants. Figure 1 presents our study design. We designed the course to support graduate students—primarily master's and Ph.D. degree candidates—in developing scientific documents, including research articles, qualification exams, and theses. The research participants comprised nine graduate students enrolled in the Scientific Writing course Of these nine participants, eight were master's degree students, and one was a doctoral student, all actively engaged in producing scientific documents as part of their graduate research requirements. All participants provided informed consent prior to participation, consistent with ethical research practices in educational settings.

Our study was conducted within an authentic graduate course setting at a single

[1]This research was submitted and approved by the Ethical Committee of the organization, and all participants signed a Free and Informed Consent Form (TCLE).

institution, limiting the participant pool to enrolled students. This context-specific approach, while constraining sample size, offers significant advantages, capturing genuine student experiences in their actual learning environment rather than in artificial experimental conditions [Creswell and Clark 2017].

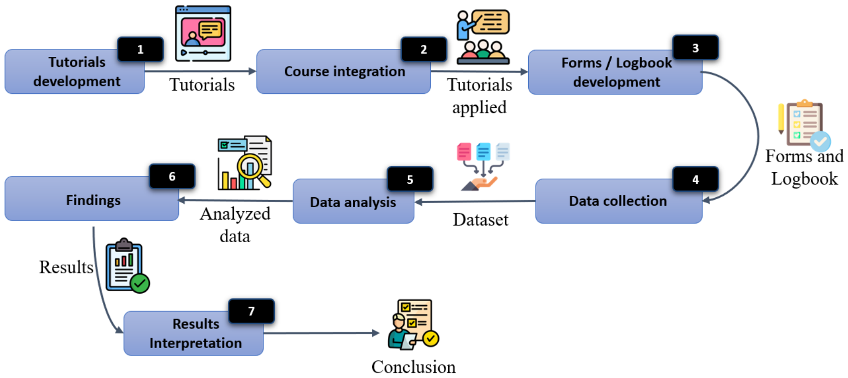

Figure 1. Research methodology: Tutorial Development, Course Integration, Data Collection, Analysis, and Interpretation

This study adopts a mixed-methods research design, combining quantitative and qualitative approaches to investigate graduate students' perceptions of GenAI tool use via designed tutorials for supporting scientific writing. Mixed-methods research allows researchers to capture both broad trends and deeper contextual understanding [Ketsman et al. 2025, Ngulube 2019].

3.1. Development of the Tutorials

We developed eleven tutorials to address critical components of scientific writing. The tutorials covered the following topics: (1) Topic Selection, (2) Scope/Problem Statement, (3) Literature Review, (4) References and Citations, (5) Scientific Methodology, (6) Experiments, (7) Analysis of Results, (8) Introduction, (9) Conclusion, (10) Abstract, and (11) Spelling and Grammar Review. These topics were selected in accordance with the IMRaD structure of scientific documents [Sollaci and Pereira 2004, Wu 2011] and with common difficulties reported in the literature regarding graduate student writing.

The development of tutorials leveraged multiple GenAI models and research software tools to ensure comprehensive, accurate, and pedagogically sound content. To facilitate hands-on learning, the tutorials direct the use of major GenAI tools (ChatGPT, Claude, and Gemini) as operational frameworks for generating explanations, examples, and structural guidance across the writing modules. Additionally, specialized AI-powered research tools were integrated into tutorial content, including: Elicit[2] for literature search and evidence synthesis; Scite.ai[3] for citation analysis and verification of research claims; Connected Papers[4] for mapping research landscapes and identifying seminal works; Perplexity[5] for fact-checking and information retrieval; ResearchRabbit[6] for discovering re-

[2] https://elicit.com/
[3] https://scite.ai/
[4] https://www.connectedpapers.com/
[5] https://www.perplexity.ai/
[6] https://www.researchrabbit.ai/

levant literature and building citation networks; NotebookLM[7] (Google) for organizing and synthesizing research materials, and others.

The tutorials were carefully designed to integrate these diverse tools, enabling the development of theoretical concepts and, especially, the demonstration of practical workflows for using AI assistance in scientific writing tasks. The original tutorials were designed to provide a step-by-step guide to approaching that particular section and to demonstrate how to effectively use GenAI tools as writing aids.

3.2. Data Collection Procedures

Data collection followed a systematic, iterative process integrated into the regular course structure throughout the semester with nine students ($n = 9$). The procedures employed a concurrent triangulation design, collecting both quantitative and qualitative data simultaneously to provide complementary perspectives on students' experiences in exploring the tutorials' content and the GenAI tools [Creswell and Clark 2017, Zhou et al. 2024].

During each class session of the Scientific Writing course, the instructor delivered a lecture on a specific topic related to scientific writing, such as methodology design, literature review strategies, or results presentation. Following the lecture, students were directed to consult the corresponding developed tutorials as they applied the concepts to their own research documents for the qualifying examination. After each tutorial, students completed a standardized questionnaire to collect quantitative data on their perceptions of the tutorial. The questionnaire included fourteen items organized into distinct thematic domains.

In addition, students maintained logbooks documenting their experiences using each tutorial and the course assignments. These logbooks served as qualitative instruments, capturing contextualized accounts of how students interacted with the tutorials, the specific challenges they encountered, the strategies they employed, and the insights they gained. Students were instructed to provide concrete examples of how they applied tutorial guidance to their writing, reflect on which aspects they found most and least helpful, and suggest improvements for future iterations.

While the questionnaire data provided breadth and enabled systematic comparison, the logbook data provided depth, revealing nuances in student experiences that standardized instruments might not capture.

The iterative nature of data collection, repeated across eleven tutorials throughout the semester, enabled a longitudinal examination of how students' perceptions and experiences evolved as they became more familiar with GenAI tools for scientific writing. All data collection procedures adhered to ethical research standards, including informed consent, voluntary participation, and assurance of confidentiality, as approved by the university's research ethics committee.

3.3. Data Analysis Procedures

Quantitative Instrument. The quantitative instrument consisted of a structured questionnaire administered via Google Forms after each tutorial. The questionnaire primarily employed 5-point Likert scales, a widely adopted format in educational re-

[7] `https://notebooklm.google/`

search that enables systematic measurement of attitudes, perceptions, and experiences [Kusmaryono et al. 2022]. The questionnaire structure aligns with established best practices in educational assessment instrument design, balancing standardized measurement through Likert scales with opportunities for detailed qualitative feedback [Johnson and Christensen 2024]. The combination of quantitative rating scales and open-ended questions enables comprehensive evaluation while maintaining efficient data collection [Creswell and Clark 2017]. This questionnaire was developed specifically for this study. No prior content or internal consistency checks were performed. The structure of the data collection form is available in Appendix A.

Qualitative Instrument. The logbook was a structured qualitative instrument designed to capture detailed, contextually rich accounts of students' experiences with GenAI tools, informed by the tutorials' content. Students were prompted to respond to five open-ended questions (as a suggestion to guide their reflection process) that systematically explored different dimensions of their interaction with the assignments, tutorials, and GenAI tools:

- *How did you use generative AI?* — This question captured students' specific workflows, strategies, and patterns of tool usage, documenting the practical ways they integrated AI assistance into their writing process.
- *How much did GenAI help you achieve the desired results in the target assignment*? — This question assessed the perceived effectiveness of AI support to evaluate the extent to which the tutorials and GenAI tools facilitated task completion and improved their writing outcomes.
- *What were your difficulties?* — This question identified challenges, obstacles, and frustrations students encountered, revealing limitations of both the tutorials and the AI tools themselves.
- *What benefits did you perceive in your experience?* — This question elicited positive outcomes, advantages, and valuable aspects of using the tutorials, documenting what students found most helpful.
- *What limitations and risks did you perceive?* — This question evaluates drawbacks, concerns, and risks associated with AI-assisted writing.

These five questions were designed to be sufficiently open-ended to allow diverse responses while providing enough structure to ensure that all participants addressed comparable dimensions of their experience, facilitating systematic cross-case analysis.

We conducted the qualitative analysis organized in three interrelated stages: **unitarization**, **categorization**, and **integrative synthesis**.

Unitarization. It corresponds to the process of deconstructing the *corpus*, in which the researcher fragments the text into units of meaning. Units of meaning are excerpts of the text (sentences, paragraphs, or smaller segments) that present their own meaning in relation to what is being investigated. This stage requires careful and repeated reading of the *corpus*, intentional selection guided by the research focus, and results in a set of interpretable fragments.

Categorization. After fragmentation, the process of reorganization begins, referred to as categorization. At this stage, the units of meaning are grouped by semantic similarity, forming broader interpretative sets. In our case, two categories defined *a priori* were used — 1) perceived benefits and 2) perceived risks and limitations — which align with the guiding questions of the student's logbook.

Integrative synthesis. The final stage is the integrative synthesis, in which the researcher reconstructs the whole based on the developed categories, producing an understanding of the object of study. In our case, the integrative synthesis was developed by considering respondents' statements across the two analysis categories.

4. Results

A total of 48 questionnaire responses were collected across the eleven tutorials implemented throughout the Scientific Writing course. The distribution of evaluations across tutorials was as follows: Topic Selection (n = 7), Scope/Problem Statement (n = 6), Literature Review (n = 3), References/Citations (n = 2), Scientific Methodology (n = 4), Experiments (n = 5), Results Analysis (n = 3), Introduction (n = 4), Conclusion (n = 3), Abstract (n = 3), and Spelling Review (n = 8). The present study (48 evaluations combined with logbook data) enables comprehensive thematic analysis while maintaining the depth characteristic of qualitative inquiry. All data from this research, including tutorials, activities, and students' answers, are available at REDU [Polastri et al. 2026].

The sample size of 48 evaluations from nine participants is methodologically justified by the concept of information power and data adequacy rather than purely statistical considerations [Malterud et al. 2016, Vasileiou et al. 2018]. In mixed-methods educational research, qualitative validity relies on the data's capacity to provide a rich, nuanced account of the phenomenon. Furthermore, the repeated-measures design—with participants evaluating multiple tutorials—effectively multiplies the analytical power, enabling robust within-subject comparisons across topics and the assessment of perceptual consistency over time [Tashakkori et al. 2020].

4.1. Quantitative Data Analysis Results

The quantitative analysis of questionnaire responses reveals that students generally have positive perceptions of the executed tutorials across multiple evaluative dimensions. We present descriptive statistics and interpretative analysis for each assessed dimension. In this evaluation, for each dimension, Likert scale questions (Items 2-9 of the questionnaire), we calculated frequency analysis, mean, median, standard deviation (SD), data dispersion, and percentages. The data analysis and synthesis of the findings were produced with the aid of GenAI.

Students' perceptions of tutorial utility (Item 2: "How useful do you think this tutorial was in supporting any aspect of scientific writing?") demonstrated strong positive ratings, with a mean score of 4.23 ($SD = 0.82$) and a median of 4.0 on the 5-point scale. The distribution showed that 35 responses (72.9$\%$) rated tutorials as 4 or 5 (useful to very useful), while only 4 responses (8.3$\%$) rated them below 3.

Similarly, the assessment of tutorial contribution to text development (Item 3) yielded a mean of 3.94 ($SD = 0.91$) and median of 4.0, with 31 responses (64.6$\%$) indicating that tutorials helped (rating 4) or helped a lot (rating 5). The learning impact dimension (Item 4) produced comparable results with a mean of 3.90 (SD = 0.94).

The overall quality of the tutorials (Item 5) received favorable ratings, with an mean of 4.21 (SD = 0.89) and median of 4.0. Notably, 37 responses (77.1$\%$) rated tutorial quality as 4 or 5 (good to excellent), with 2 responses (4.2$\%$) providing ratings below 3.

The assessment of whether the tutorials accelerated the scientific writing process (Item 6) yielded a mean of 3.98 ($SD = 1.12$) and median of 4.0, indicating moderately positive perceptions, but with greater variability than on other dimensions. Although 32 responses (66.7%) indicated acceleration (scores 4–5), 8 responses (16.7%) rated this dimension with a score of 3 (neutral), and 8 responses (16.7%) indicated minimal acceleration (scores 1–2).

The clarity and organizational structure of the tutorial (Item 7) received positive evaluations, with an average of 4.31 (SD = 0.92), the highest among all quantitative dimensions, and median of 5.0. Forty responses (83.3%) rated clarity as 4 or 5, whereas 2 responses (4.2%) rated it 3 or lower.

The dimension that assesses the contribution to understanding how to apply GenAI in scientific writing (Item 8) had an average of 4.10 (SD = 0.88) and median of 4.0, with 36 responses (75.0%) indicating a substantial contribution (scores 4-5).

Students' perception of the impact of the tutorials on improving writing quality, such as clarity, coherence, and style (Item 9), averaged 3.94 (SD = 1.06) and median of 4.0. Although 33 responses (68.8%) indicated a positive impact (scores 4-5), this dimension showed slightly greater variability than the others, with 7 responses (14.6%) being neutral and 8 responses (16.7%) indicating minimal impact.

Responses to the dichotomous question regarding technical or conceptual difficulties (Item 10) revealed that 21 participants (43.8%) reported no difficulties, 20 (41.7%) reported partial difficulties, and 7 (14.6%) reported difficulties.

The assessment of ethical preparedness following tutorial use (Item 11) yielded positive outcomes, with 35 responses (72.9%) indicating feeling prepared, 12 (25.0%) indicating partial preparedness, and only 1 (2.1%) indicating not prepared. Similarly, regarding whether tutorials encouraged critical reflection and avoided excessive dependence (Item 12), 32 responses (66.7%) indicated positive encouragement, 11 (22.9%) indicated partial encouragement, and 5 (10.4%) indicated negative encouragement.

Across all quantitative dimensions, median values ranged from 4.0 to 5.0, with seven of eight dimensions showing a median of 4.0 and one dimension (Item 7: Clarity and Structure) achieving a median of 5.0. The consistency between median and mean values across most dimensions strengthens confidence in our findings, suggesting that positive perceptions are not driven by outliers but represent genuine trends across the student group. Notably, the median of 5.0 for clarity and structure indicates that more than half of all students rated this dimension at the maximum level, further validating the instructional design approach employed in tutorial development.

4.2. Qualitative Data Analysis Results

The written responses in the students' logbooks were grouped During the unitization and categorization stages, excerpts from the two analysis categories were identified: **perceived benefits** (category 1) and **perceived risks and limitations** (category 2). This process was carried out independently by a single researcher, given the clarity of the analytical categories. Table 1 (in Appendix B) shows a summary of the benefits, as well as the perceived risks and limitations.

One of the most common benefits is support for searching and selecting references. Students cited using tools such as *Connected Papers* and *Elicit* to map relevant articles and identify gaps in the literature. One participant reported: "*I used GenAI to find better articles for references in the text sections. It helped a lot, since it was used to improve the text's references*" (free translation by the authors). Another participant reported using Gemini to "converse with some denser articles", showing that they used AI to assist in understanding complex texts.

Another frequently mentioned aspect concerns the generation of textual structures and skeletons. AI was used to propose section divisions, suggest subheadings, and organize topics. As reported: "*I used language models to generate text skeletons, which facilitate writing and organizing ideas for me*".

In the context of writing, grammatical revision, style refinement, and improved textual fluency appear as benefits. AI was employed to "correct and improve the connection between different ideas", and to "revise passages and improve methodological clarity". Brainstorming and the exploration of alternatives also emerged as potential strengths. In some cases, AI even suggested methodological adjustments: "*it suggested scope changes that were considered relevant and adopted to readjust the schedule*".

Time optimization is perhaps the most transversal benefit among the reports. Expressions such as "considerably sped up the work", "time savings", and "greatly accelerated the writing process" appear frequently. One student observed that the tool enabled them to "focus more on conceptual decisions than on language adjustments", suggesting that the time saved was spent on the project's analytical and methodological dimensions.

Regarding **perceived limitations and risks**, the most evident theme in the responses refers to the veracity of the information. As several participants summarized, the main concern was "verifying whether the content or information was truthful". Another student reinforced that "the main risk is trusting suggestions without critically analyzing what really needs to be done in the project".

The reliability of the extracted data was highlighted as a critical issue, often requiring additional work for the researcher. One participant reports that, after manual verification, "at least 25% of the information extracted by the platform was incorrect or misinterpreted and required correction". The issue of bibliographic references also arose. One student reports that "*the LLM's ability to work with citations is extremely deficient: it either generates errors or includes citations that, although correct, are very tangential to the topic in question*".

Regarding text quality, criticisms arise concerning repetitiveness and verbosity. Some students report: "*When longer texts are generated, they are always very repetitive and have a scope that is very limited to the prompt used*". Another limitation refers to the tool's difficulty in maintaining scope. One participant pointed out that the system "*sometimes fails to stay within the specified scope, even when limitations are manually defined and emphasized in the prompt*".

In the field of bibliographic search, a student highlighted the limitations of database access. He emphasizes that "when performing searches, ChatGPT cannot access paid or restricted access articles", leaving out repositories such as IEEE and Scopus. Concerns also emerged regarding methodological rigor in systematic reviews. One student stated: "*I chose not to use GenAI in the screening phase of the articles. This step was perfor-*

med manually to ensure greater control and methodological rigor in the application of inclusion and exclusion criteria".

5. Discussion

The quantitative analysis demonstrated that students perceive the tutorials as highly useful educational resources. The strongest ratings were observed for clarity and structure (M = 4.31, SD = 0.92), perceived utility (M = 4.23, SD = 0.82), and overall quality (M = 4.21, SD = 0.89), with over 70% of responses rating these dimensions as good to excellent. These findings align with previous research [Kasneci et al. 2023, Daza et al. 2024] indicating that GenAI can provide experiences that enhance student engagement.

The robustness of these findings is further supported by median values, which showed remarkable consistency at 4.0 across seven dimensions and 5.0 for clarity and structure. The alignment between median and mean values indicates that positive perceptions are not artifacts of a few extremely favorable responses but reflect consistent trends across the student cohort.

Variability emerged in dimensions related to process acceleration (M = 3.98, SD = 1.12) and writing improvement (M = 3.94, SD = 1.06), suggesting that the perceived efficiency gains and quality enhancement differ substantially across individual students. This variability may reflect differences in prior familiarity with GenAI tools, individual writing processes, or the specific nature of writing tasks addressed by different tutorials.

The qualitative analysis revealed a nuanced understanding of AI's role in scientific writing. Students consistently positioned GenAI as an "assistant" rather than a replacement for human authorship, corroborating recent findings on hybrid authorship in academic writing. The most valued benefits included support for reference identification and qualification, text structuring, time optimization, and brainstorming. Students particularly appreciated AI's capacity to help them "focus more on conceptual decisions than on language adjustments", suggesting that these tools can redirect cognitive resources toward higher-order analytical tasks.

Conversely, students identified critical limitations that warrant serious pedagogical attention. The most prominent concern involved information veracity, with one student reporting that "at least 25% of the information extracted by the platform was incorrect or misinterpreted and required correction". This finding is consistent with documented reports of "hallucinations" in LLMs [Granjeiro et al. 2025, Ren and Wang 2025]. Additional concerns included repetitive outputs, difficulty maintaining specified scope, limited access to paywalled databases, and deficient citation capabilities—issues that align with the "reference validity gap" identified in prior research [Ren and Wang 2025].

Approximately 72.9% of students reported feeling ethically prepared to use AI responsibly after using the tutorials, and 66.7% indicated that the tutorials encouraged critical reflection. However, the remaining percentages suggest that not all students feel ethically prepared for AI use, underscoring the continued relevance of explicit ethical guidance in AI-integrated curricula.

These findings carry significant implications for integrating GenAI into graduate education. First, the high ratings for clarity, structure, and utility suggest that well-designed tutorials can effectively complement traditional instruction. However, the moderate ratings for process acceleration and writing improvement indicate that educators should set realistic expectations about what these tools can achieve. Second, the qualitative findings underscored the relevance of teaching verification and validation skills alongside AI literacy. This aligns with recommendations that AI support should be combined with instructor guidance to achieve optimal educational outcomes [Kasneci et al. 2023].

Third, the identification of AI as an “assistant”rather than a substitute reflects awareness of appropriate roles for AI. Educators should reinforce this positioning while remaining vigilant for signs of over-reliance or “cognitive debt”—the phenomenon wherein immediate efficiency gains come at the cost of reduced critical engagement with content [Kosmyna et al. 2025]. Fourth, the time optimization benefits reported by students suggest that such designed tutorials can help address the documented burden on both students and supervisors in graduate writing instruction [Gupta et al. 2022]. This efficiency should be channeled toward deeper engagement with the research substance.

Our findings revealed positive perceptions and identified relevant challenges and areas for improvement. From a broader perspective, the adoption of GenAI to support scientific writing must also be understood within its social context. The increasing integration of AI systems into academic practices reshapes how knowledge is produced, evaluated, and disseminated, raising important sociotechnical questions about authorship, intellectual autonomy, and epistemic responsibility. In this sense, tools that assist in the construction of scientific texts do not merely automate writing tasks; they participate in the evolving social organization of scholarly work. Reflecting on their use, therefore, requires considering not only technical effectiveness but also their implications for academic culture, power relations in knowledge production, and the equitable development of research skills across different communities. Unfortunately, the distribution of responses across the tutorials was highly asymmetrical. We agree that the consolidated results can serve as initial exploratory evidence, not for broad generalizations.

6. Conclusion

This study investigated graduate students’ perceptions of using GenAI tools for scientific writing through guided tutorials in a Scientific Writing course at the Institute of Computing of the University of Campinas. Results showed that students positively evaluated the tutorials and perceived GenAI as useful for structuring texts, brainstorming, identifying references, and saving time. However, concerns regarding information accuracy, citation reliability, repetitive outputs, and limited access to paywalled sources highlighted the importance of critical validation. Our findings provide empirical evidence on the benefits and challenges of integrating GenAI into graduate education and reinforce the need for pedagogical guidance, ethical awareness, and verification skills. Overall, tutorials show potential to support scientific writing, although effective adoption requires careful integration and further investigation of long-term educational impacts.

Acknowledgements

This work was supported by the National Council for Scientific and Technological Development (CNPq), Brazil, under grants #301337/2025-0 and #400456/2025-7.

Referências

Bhutoria, A. (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence*, 3:100068.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Cahill, L., Miller-Cochran, S., Pantoja, V., and Rodrigo, R. L. (2008). Graduate student writing groups as peer mentoring communities. *Stories of mentoring: Theory and praxis*, pages 153–58.

Creswell, J. W. and Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Daza, V., Martí-Testón, A., Casañ, M. J., and Castelló-Cogollos, L. (2024). Student perceptions of generative artificial intelligence in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 21(1):52.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). Opinion paper: ”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.

Fanguy, M., Costley, J., Baldwin, M., Lange, C., and Wang, H. (2019). Diversity in STEM: Influence of demographic characteristics on learning preferences and educational technology use. *Journal of Science Education and Technology*, 28(6):556–568.

Gopee, N. and Deane, M. (2013). Strategies for successful academic writing: Institutional and non-institutional support for students. *Nurse Education Today*, 33(12):1624–1631.

Granjeiro, J. M., Cury, A. A. D. B., Cury, J. A., Bueno, M., Sousa-Neto, M. D., and Estrela, C. (2025). The future of scientific writing: Ai tools, benefits, and ethical implications. *Brazilian Dental Journal*, 36:e25–6471.

Gupta, S., Jaiswal, A., Paramasivam, A., and Kotecha, J. (2022). Academic writing challenges and supports: Perspectives of international doctoral students and their supervisors. In *Frontiers in education*, volume 7, page 891534. Frontiers Media SA.

Hwang, G.-J. (2014). Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments*, 1(1):1–14.

Imran, M. and Almusharraf, N. (2023). Analyzing the role of chatgpt as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4):ep464.

Inouye, K. and McAlpine, L. (2019). Developing academic identity: A review of the literature on doctoral writing and feedback. *International Journal of Doctoral Studies*, 14:1–31.

Johnson, R. B. and Christensen, L. B. (2024). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage publications.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274.

Ketsman, O., Droog, A., and Qazi, S. (2025). Mapping the prevalence of mixed methods research in educational technology journals. *Computers & education*, 226:105207.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., and Maes, P. (2025). Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4.

Kusmaryono, I., Wijayanti, D., and Maharani, H. R. (2022). Number of response options, reliability, validity, and potential bias in the use of the likert scale education and social science research: A literature review. *International Journal of Educational Methodology*, 8(4):625–637.

Lax, J. R. (2014). Engaging graduate students in the practice of authoring scientific texts. *International Journal of Science Education*, 36(10):1636–1662.

Lendvai, G. F. (2025). Chatgpt in academic writing: A scientometric analysis of literature published between 2022 and 2023. *Journal of Empirical Research on Human Research Ethics*, 20(3):131–148.

Lin, L. H. and Morrison, B. (2021). Challenges in academic writing: Perspectives of engineering faculty and l2 postgraduate research students. *English for Specific Purposes*, 63:59–70.

Lo, C. K. (2023). What is the impact of chatgpt on education? a rapid review of the literature. *Education sciences*, 13(4):410.

Lonka, K., Hietajärvi, L., Hakkarainen, K., Munter, T., Palonen, T., and Salmela-Aro, K. (2019). The challenges of doctoral writing: Experiences of Finnish PhD students. *Scandinavian Journal of Educational Research*, 63(7):1019–1036.

Lu, X. and Wang, C. (2022). A systematic review of students' self-regulated learning in technology-enhanced learning environments. *Education and Information Technologies*, 27(2):1825–1862.

Luitse, D. and Denkena, W. (2021). The great transformer: Examining the role of large language models in the political economy of ai. *Big Data & Society*, 8(2):20539517211047734.

Maher, M. A., Timmerman, B. C., Feldon, D. F., and Strickland, D. (2014). Factors affecting the occurrence of faculty-doctoral student co-authorship. *The Journal of Higher Education*, 85(1):121–149.

Malterud, K., Siersma, V. D., and Guassora, A. D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative health research*, 26(13):1753–1760.

McAlpine, L. and Amundsen, C., editors (2011). *Doctoral education: Research-based strategies for doctoral students, supervisors and administrators*. Springer Science & Business Media.

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., et al. (2013). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine*, 89(2):236–243.

McRell, A. S., Wilson, B. L., and Levkoff, S. E. (2021). Writing skills development for graduate studies and career readiness in science and aging fields: A case study approach. *Frontiers in public health*, 9:727064.

Mienye, I. D., Jere, N., Obaido, G., Ogunruku, O. O., Esenogho, E., and Modisane, C. (2025). Large language models: an overview of foundational architectures, recent trends, and a new taxonomy. *Discover Applied Sciences*, 7(9):1027.

Ngulube, P. (2019). Mapping methodological issues in knowledge management research, 2009–2014. *International Journal of Knowledge Management (IJKM)*, 15(1):85–100.

North, S. (2005). Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education*, 30(5):517–533.

Odena, O. and Burgess, H. (2017). How doctoral students and graduates describe facilitating experiences and strategies for their thesis writing learning process: A qualitative approach. *Studies in Higher Education*, 42(3):572–590.

Peláez-Sánchez, S., Jordano, M., Freire, T., Verdugo-Castro, S., and Sánchez-Gómez, M. C. (2024). Perceptions of university students on the use of ChatGPT in education. *Education Sciences*, 14(1):94.

Peña-Lima, C., Dildine, T. C., and Lee, R. (2026). Three tips for scientific writing: a guide for graduate students. *Nature*.

Polastri, P. C., Linhalis, F., and dos Reis, J. C. (2026). Conjunto de dados e materiais pedagógicos sobre o uso de inteligência artificial generativa na escrita científica para pós-graduação. Repositório de Dados de Pesquisa da Unicamp (REDU). Disponível em: https://doi.org/10.25824/redu/BLSSTA.

Pololi, L. H., Knight, S. M., Dennis, K., and Frankel, R. M. (2011). Helping medical school faculty realize their dreams: An innovative, collaborative mentoring program. *Academic Medicine*, 77(5):377–384.

Ren, J. and Wang, W. (2025). Assisting research proposal writing with large language models: Evaluation and refinement. *arXiv preprint arXiv:2509.09709*.

Scalfani, V. F., Maceli, K. M., Towell, J. L., Turner, J. A., and Williams, A. E. (2016). Filling the gap: A program to assist graduate students with research and scholarly publication. *portal: Libraries and the Academy*, 16(1):125–145.

Shah, J., Shah, A., and Pietrobon, R. (2009). Scientific writing of novice researchers: what difficulties and encouragements do they encounter? *Academic Medicine*, 84(4):511–516.

Shopovski, J., Mohdali, R., and Marolov, D. (2025). Revolutionizing peer review: A comparative analysis of chatgpt and human review reports in scientific publishing.

Sollaci, L. B. and Pereira, M. G. (2004). The introduction, methods, results, and discussion (imrad) structure: a fifty-year survey. *Journal of the medical library association*, 92(3):364.

Song, C. and 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.

Su, J. and Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3):355–366.

Sullivan, M., Kelly, A., and McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 6(1).

Tashakkori, A., Johnson, R. B., and Teddlie, C. (2020). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage publications.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and 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(1):15.

van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., and Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, 614(7947):224–226.

Vasileiou, K., Barnett, J., Thorpe, S., and Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC medical research methodology*, 18(1):148.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wu, J. (2011). Improving the writing of research papers: Imrad and beyond. *Landscape Ecology*, 26(10):1345–1349.

Zhou, Y., Zhou, Y., and Machtmes, K. (2024). Mixed methods integration strategies used in education: A systematic review. *Methodological Innovations*, 17(1):41–49.

Appendix A

Data Collection Form Structure

The questionnaire comprised fourteen items organized into distinct thematic domains. The first item asked students to identify which tutorial they were evaluating, selecting from the eleven available options. The questionnaire evaluated:

- Perceived utility (Item 2): "How useful do you think this tutorial was in supporting any aspect of scientific writing?" (1 = Not useful at all, 5 = Very useful);
- Support for text development (Item 3): "How did this tutorial help in developing the text?" (1 = Did not help, 5 = Helped a lot);
- Learning impact (Item 4): "How much did this tutorial help your learning?" (1 = Did not help, 5 = Helped a lot);
- Tutorial quality (Item 5): "How do you evaluate the quality of this tutorial?" (1 = Terrible, 5 = Excellent);
- Process acceleration (Item 6): "Did using this tutorial accelerate the scientific writing process?" (1 = Did not accelerate, 5 = Accelerated a lot);
- Clarity and structure (Item 7): "How do you evaluate the clarity of instructions and organization of this tutorial?" (1 = Very confusing, 5 = Very clear);
- Learning and applicability (Item 8): "How much did this tutorial contribute to your understanding of how to apply generative AI in scientific writing?" (1 = Did not contribute at all, 5 = Contributed completely); and
- Writing improvement (Item 9): "To what extent did using the AI tools suggested in this tutorial help you improve aspects such as clarity, coherence, and text style?" (1 = Did not impact, 5 = Impacted completely).

Three dichotomous and trichotomous items assessed specific concerns related to AI use in academic contexts:

- Technical difficulties (Item 10): "Did you encounter technical or conceptual difficulties using the AI tools demonstrated in this tutorial?" (Yes / Partially / No);
- Ethical preparedness (Item 11): "After using this tutorial, do you feel more prepared to use AI tools ethically and responsibly in scientific research?" (Yes / Partially / No); and
- Critical autonomy (Item 12): "Did this tutorial encourage critical reflection on when and how to use AI in producing scientific texts, avoiding excessive dependence?" (Yes / Partially / No).

Finally, two open-ended questions enabled students to provide qualitative feedback:

- Areas for improvement (Item 13): "Which aspects of this tutorial do you believe could be improved (for example: practical examples, pace, language, interface, scientific contextualization)?" and
- Overall experience (Item 14): "How would you describe your overall experience using this tutorial on generative AI in scientific writing?"

Appendix B

Table 1. Summary of perceived benefits and risks (our findings) from the qualitative data analysis.

Perceived Benefits	**Perceived Risks and Limitations**
– Identification of references;	– Lack of reliability (need for constant validation);
– Text structuring;	– Repetitive responses;
– Linguistic revision;	– Difficulty staying within scope (drifting from the prompt);
– Brainstorming;	– Limited access to private databases;
– Methodological organization;	– Lack of methodological rigor.
– Time optimization.	