Human-in-the-Loop Writing: Students Self-Assessing Their Use of GenAI Applications

Josh Barrows, Maryam Vaezi, and Rochelle Rodrigo, University of Arizona

This study aims to shed light on the composition students' perceptions and interactions with generative AI (GenAI) tools integrated into writing instruction. While the small sample size limits generalizability, structured reflection revealed nuanced student attitudes about the process, efficiency, ethics, and quality of artificial intelligence (AI)-assisted writing. Student evaluations of AI tools involved several key dimensions: process, efficiency, ethics, and quality. Students assessed how well these tools supported specific writing process stages, including reading, grammar, summarizing, analyzing, outlining, brainstorming, and source finding. Efficiency emerged as a significant factor in students' evaluations, specifically ease of use and speed. Ethical considerations also played a crucial role, with students expressing varied levels of confidence and uncertainty about issues related to academic integrity, plagiarism, and the responsible use of GenAI. The quality of the AI output was a critical focus, with students assessing its effectiveness based on prompt responsiveness, detail, adherence to genre conventions, and comparative evaluations against their own writing or the results from other Gen AI applications.

The recent advancements in Large Language Models (LLMs), including ChatGPT and other generative artificial intelligence (GenAI) tools, have prompted a significant increase in scholarly research addressing their applications, challenges, and ethical considerations in higher education (Kostopolus, 2025). While digital literacy is formally embedded in many curricula, the success of technology integration ultimately depends on whether teachers and students personally choose to adopt these tools (Barron et al., 2003). Adoption is a complex social and developmental process shaped by interacting factors such as personal traits, prior experiences, the nature of the innovation, and contextual influences that evolve over time (Adler & Clark, 1991; Hall, 1979; Rogers, 1995). Students' willingness to adopt artificial intelligence (AI) tools like ChatGPT is significantly influenced by their perceptions of usefulness, ease of use, emotional responses, and past experiences (Hord et al., 1987; Zou & Huang, 2023).

This study was developed to collect data about students' perceptions of using GenAI applications in writing courses. In this chapter we share codes

 DOI: https://doi.org/10.37514/PCW-B.2026.2883.2.05

developed from students' reflection on the efficacy of being assigned and using Gen AI applications. While the small sample size limits generalizability, structured reflection revealed nuanced student attitudes about the process, efficiency, ethics, and quality of AI-assisted writing.

Not surprisingly, research already shows there are many reasons students adopt using GenAI applications. Al-Mamary and Abubakar (2025) investigated the adoption of ChatGPT as an AI tool by university students in Saudi Arabia, and "perceived autonomy and relatedness" (p. 1) were found to be the key factors influencing students' willingness to embrace AI technologies, while perceived competence did not significantly affect this willingness. In a similar vein, Strzelecki (2024) examined the factors influencing the adoption of ChatGPT by 503 students from Polish state universities. Several key factors were identified: a) "hedonic motivation", which refers to the enjoyment and satisfaction students derive from using AI and its subsequent effect on their motivation to use it, b) "performance expectancy", relating to students' perceptions of ChatGPT as a means to enhance their academic performance, and c) "habit", reflecting the regular integration of AI tools, including ChatGPT, into students' study routines (p. 237). These findings are consistent with Segreeva et al.'s (2025) study, which identified close links between "habit", "performance expectancy, "and "hedonic motivation" and the students' acceptance and use of AI tools in higher education (p. 223). Segreeva et al. (2025) also introduced "social influence" (SI), which includes peer encouragement, and "price value" (the perception of affordability) as significant predictors of students' "behavioral intention (BI)" to adopt these technologies (p. 227).

Over the past few years, researchers have found students use GenAI applications for a variety of reasons. Črček and Patekar's (2023) study on students' use of ChatGPT for academic writing in Croatian public and private universities revealed that more than half of the students had used ChatGPT, primarily for generating ideas, paraphrasing, and summarizing content. Črček and Patekar found that while many students viewed using ChatGPT to generate ideas as ethically acceptable, fewer considered it ethical to use the tool for other purposes or to replace their own writing. Surveying students across various disciplines in a Saudi Arabian university, Bensalem et al. (2024) found that most participants regularly use tools like Grammarly, ChatGPT, and ProWritingAid for such benefits as improved grammar, coherence, vocabulary, and productivity. Despite the generally positive perception of these tools, some students expressed concerns about becoming overly dependent on them, the potential negative effects on creativity and critical thinking, and ethical issues related to AI-assisted writing.

Students' feelings about using GenAI are just as varied as their reasons for adoption and use. Maya's (2024) work highlights a wide range of student

attitudes towards AI. While most students used AI for efficiency, stress relief, and personalized support, Maya found others also avoided it. The diversity of student attitudes toward AI is consistent with findings from Cummings et al.'s (2024) study, which examined first-year writing courses incorporating GenAI tools. According to Cummings et al., students valued the support these tools offered but expressed concerns about losing their authorial voice, becoming overly dependent, and potentially committing plagiarism. Structured assignments and reflective practices helped students overcome these concerns and critically assess the tools' usefulness. Consequently, most students started to see AI as a helpful supplement rather than a replacement (Cummings et al., 2024). Like Maya (2024), Bensalem et al. (2024) and other scholars, Cummings et al. (2024) pointed out growing yet ambivalent adoption of AI tools in student writing, highlighting concerns about ethics, dependence, and voice alongside recognized benefits.

In their empirical study involving 603 students and 312 writing instructors, Gallagher and Wagner (2024) compared student and instructor perceptions of academic dishonesty in collaborative writing situations involving humans vs. AI writing technologies. Findings revealed that students and instructors generally aligned in their perceptions, with both groups demonstrating "significant differences in perceived dishonesty between AI and human collaborators" in certain cases (Gallagher & Wagner, 2024, p. 267). Both teachers and students perceived AI to be more academically dishonest than human collaborators in most contexts, particularly when AI was responsible for generating the text. However, both groups associated less academic dishonesty with AI when it was used for lower-level processes such as brainstorming and language refinement. Findings also reveal a shared uncertainty among participants. For instance, both students and instructors appeared unsure about whether having a human or an AI tool generate an outline should be considered academically dishonest. Gallagher and Wagner (2024) note that this uncertainty highlights the importance of tackling questions of academic integrity in genuine, real-world contexts.

Methods

Our Intuitional Review Board (IRB) approved study focused on students' use of GenAI applications in their writing courses. The research team designed their writing courses to incorporate GenAI applications in the teaching, learning, and writing processes. One course, ENGL102 (a second semester first-year composition [FYC] course), was taught by Barrows and Rodrigo. The second course, ENGL307 (Introduction to Business Writing), was taught by Barrows. As part of each course, students were prompted to reflect upon

their expectations and experiences with the GenAI applications. Specifically, students were prompted to:

1. Complete an initial survey that asks qualitative and quantitative questions about previous experiences with writing, online learning, and GenAI applications.
2. Participate in an initial discussion board activity that asks students to a) reflect on their pre-existing beliefs and expectations about GenAI and b) learn how GenAI works and then reflect upon what they've learned.
3. Complete a short survey after every activity or assignment that includes using GenAI.
4. Respond to assignment specific reflection questions for each of the assignments that used GenAI.
5. Produce an end-of-course reflection on their use of and learning about GenAI.
6. Complete an end-of-course survey that includes parallel questions from the initial survey.

All the "research" instruments were assigned as reflective learning assignments, which is part of the DEER Praxis that we used to help design the course materials (Cummings, Monroe, Watkins, 2024). After the course was completed and grades were posted, students were asked to provide consent for these learning materials to be used for this study.

Our research team met for four sessions where we compared our individual coding. In between these sessions, we had many two-person sessions which we used to collaboratively develop and refine the code book as well as code the data. For each major session, the third member corroborated the two independent codes to find areas of disagreement. Our final codebook includes four major sections focused on how students assessed the usefulness or efficacy of the GenAI tools. The students assessed the tools based on process, efficiency, ethics, and quality.

We had six respondents with three from each course. These participants are represented as pseudonyms, and we did not correct or revise any of the student language presented in quotes in our discussion. Due to the small sample size, these results are not generalizable; but they can provide some important information for how we understand the ways that students might be interacting with and perceiving use of GenAI in the writing classroom.

Data Analysis

We developed four codes about how students were assessing the efficacy of the GenAI tools that they used in the course: process, efficiency, ethics, and

quality. With many of our sub-codes there were instances that students articulated general codes with either positive or negative connotations. So, we further delineated between many codes by appending either a positive (p)/yes (y) or negative/no (n) with that code.

Assessing Based on Process

The *process* category of codes was based upon students' claims, or predictions, of the usefulness of general and sometimes specific GenAI applications based on how they would help with specific parts of the writing process. Within the process category we had six codes: read, grammar, summarize, analyze, outline, brainstorm, and find source.

The first code, *read*, refers to any instance where students note that GenAI has assisted in their reading or understanding of the material. When asked about using GenAI in the future, Alara said "Gen AI might help me in the sense that it can help me understand the topic a little better to really be able to become as knowledgeable as I can be." Although we never explicitly prompted students to read a text through or with GenAI, it appears that assigning students to prompt for summaries functioned as ways to engage more meaningfully with the text.

The second code, *grammar*, refers to any time that students note that the AI application has assisted with grammar revision or recommendations for writing. Meredith describes how "Gen AI can also help point out grammar mistakes i didnt notice earlier." Other participants also articulated grammar as an area that GenAI can help them. We found the *grammar* comments interesting since we never explicitly assigned an activity that prompted students to use a GenAI application to analyze their drafts, especially for tone, style, and grammar. Students already brought the knowledge that reviewing drafts for tone, style, and grammar was something a GenAI application might do; and they obviously value that help.

The third process code, *summarize*, is used to mark where students note how the AI application has summarized or paraphrased input content. James reflected: "I could use gen AI to generate the summaries from the notes, and then proof read the summary and correct it where need be as well as reword it to insure that it is cohesive. I think that Gen AI in this case is a perfect application as one of the things that it does best is summarize. It also saves me some of the trouble of trying to balance important points, as I can highlight all of them and then generate a summary based on those." We find the comments on *summary* interesting, especially in the cases where students are articulating how they can improve the summary by including their own annotations from their first read through of the article in the

prompt for the AI, like James does. This type of awareness goes beyond just asking AI to summarize an article, but it is a purposeful summative task with directive elements.

The fourth process code, *analyze*, is used to show where students express how the AI application sometimes analyzes readings or analyzes their own work as a reviewer, including drafts, notes, and instructions. Alex projected future use: "I will start by addressing whatever homework precedes the step of drafting the review of literature, then create an outline of what I want it to look like. I will then formulate a draft and utilize AI to "poke holes" in what I have and receive suggestions for what to add." The *analyze* code is important because we realized that students are using the AI application to analyze both their own writing and argument like Alex does as well as using it to analyze a particular source like other students mentioned.

The fifth process code, *outline*, refers to when the student responses show that the AI application has provided an outline for input content or provides a suggested outline for student work. James describes how the AI application can do outline work, saying "In this case I was much more excited about the summaries and the RoL drafts or outlines that it made, …. This will hopefully help me in writing my RoL."

The sixth process code, *brainstorm*, refers to when students articulate how an AI application has developed new and creative ideas or offered different perspectives. Many students, like Alex, claimed "Gen AI will help me think of ideas and how I can improve upon the existing work that I have." The students expressed brainstorming both in the ideation stage of writing as well as in the revision stage, articulating how the AI application can be used as a brainstorm in multiple areas of their writing process.

Finally, the last process code, *find source*, describes how students articulate that the AI application helps students explore related resources they already know or when the application finds initial sources related to a topic. Bethany projected future use: "I plan to find sources that branch off each other to fill in the gaps the other articles are lacking." Since we never explicitly prompted students to use GenAI to find sources, the *find source* code emphasizes the need to make sure students understand how specific GenAI applications work and what information and resources to which the GenAI application has access.

Since we specifically prompted students to reflect, we believe many of these process comments are heavily influenced by the required prompting after a specific activity. We think the most honest assessment based on the writing process was when students were prompted to predict when and how they might use GenAI in future composing situations.

Assessing Based on Efficiency

The second major category that developed in our iterative coding was how students were assessing AI application efficacy based on *efficiency*. We are using two sub-codes for *efficiency*. The first code for efficiency, *easy use* (y/n), refers to instances where students specifically mention that the GenAI application was easy to use as well as times where they mentioned that it was specifically not easy to use. Alex claimed "Copilot was the worst because it would not allow me to include all of the notes I needed it to analyze in the prompt. It also didn't seem to comprehend what I was asking it to do when I told it to not respond until I explicitly told it to, nor when I asked it to address my three previous prompts." Not surprisingly, we find the *easy use* code demonstrates that the usability of the tool is one of the important measures that students are using to determine GenAI application efficacy.

The second code for *efficiency*, speed, refers to times where students articulate how quickly, or more rapidly, work is accomplished with a GenAI application. Two students also claimed the reverse; for example, Bethany reflected: "Some concerns I have from using Gen AI is having to navigate and refine questions consistently. It takes about as long as it does to refine questions in AI as it does from a general google search." These moments are helpful for instructors to encourage students to think about their writing process and ways that the GenAI application can sometimes be helpful, or faster, in some ways and not in others.

Assessing Bas ed on Ethics

Our pedagogical framework also attends to ethics in our instruction, and we found that students also expressed their perceptions of ethics as an important piece of their assessment of AI applications in the writing classroom. The third major theme of *ethics* is broken into three sub codes: confident, unsure, and stores.

The first code, *confident*, refers to both instances of when a student is confident that their use of GenAI was ethical as well as when they were confident that the use of GenAI is not ethical. Alex reflected: "I am most excited about the content that addresses specific questions, rather than the ones with specific prompts asking the AI to generate entire assignments. The reason for this is because I think it is a slippery slope to ask AI to do too much on the front end, and a student may risk plagiarism because of it."

Our second code for ethics, *unsure*, is used to notate when students specifically mention that they are unsure whether the specific use of AI is allowed as well as when the student articulates that they are unsure whether the

specific use was unethical. Meredith reflected: "My main concern I have for AI is when I ask it for help with my writing, sometimes it will give me a great sentence that I want to use, but if I use it it is plagiarism. So I will try to reword it into my words. But still, is that a form of plagiarism?" Students articulated a lot of concern about plagiarism and academic integrity when using GenAI applications. In future iterations, we might use additional codes since *allowed* and *ethical* are not synonymous.

The third ethics code, *stores*, refers to instances where students articulate how the GenAI application is used to store ideas, notes, or sources where the student can return to them. James says, "I generally don't use AI much outside of the assigned usage, but if I do it is usually in powernotes, and so I have a recorded history in the project that allows me to track everywhere I have used it and to be able to accurately recall where and when it was used." In these instances, students are outsourcing confidence in the tools to their tracking abilities or when they are being intentional about archiving material and instances of GenAI use.

Assessing Based on Quality

Finally, and perhaps most significantly, students assessed AI applications based on the quality of their performance. We determined five subcodes assessing quality: prompt, not enough, more, form, and compare.

The first code, *prompt (y/n)*, refers to instances where students assessed accuracy broadly. Students specified whether the prompt works (or not) as a measure of the quality of the result. James was not happy with summary results: "I don't think it was necessarily wrong, but it was very short to summarize the entire purpose of the article, and I think it gets at too much of the literal function of the writing and less at the bigger picture stuff like why this was written, or the purpose of studying this material at all." We typically saw the *prompt* code while students were comparing different AI tools and how some of them functioned better or worse than others at a specific writing task.

The second code in quality, *not enough*, indicates that students expressed how the AI application's generated results did not include enough information. Alex describes how "The worst in this case was Powernotes; it focused more on what the article implications are and didn't assess the audience, the exact article name or author, and other crucial details." Here, it is evident that Alex has a contextual understanding of what a summary should include and the AI application did not include it. These cases are opportunities for instructors to think about rhetorical re-prompting to guide the AI to give output according to what students are looking for.

The third quality-based code, *more (p/n)*, refers to when students describe how the AI application included more information that is sometimes good and is sometimes excessive, negative, or bad. An example of more (p) and more (n) are:

- Bethany: "I prefer to do the genre analysis with AI as the AI provided more information and ideas that I was not aware of."
- Alex: "as long as the findings are specific while concise, there is no reason to include every specific finding in the summary. The summary is meant to concisely summarize, not to re-write the introduction and conclusion."

Although almost all of our participants articulated that the GenAI applications provided more information than what they would have found on their own, a couple of the participants, like Alex, show that sometimes a more detailed response is not desired.

The fourth code, *form (y/n)*, is used to indicate when students describe how the AI application generated results that followed expected conventions or not; conventions include grammar, length (conciseness), style, and punctuation. James reflected, "I was a little surprised when I prompted the AI to write it with a policy maker audience in mind that it went straight to writing a letter pitching information to them, but it also makes a lot of sense to write it in that way, it just doesn't serve my purposes perfectly." For students to assess the form of GenAI application output, they need to understand the genre conventions for the type of text they are hoping to create.

The fifth code, *compare*, refers to when students compare the results with something else to assess quality. Sometimes students are comparing the AI application to their own work, and sometimes students are comparing to a different GenAI application or a different source. Meredith compared results from GenAI to her own work: "The difference between ours and AI is that the AI one gave a lot more detail than ours it also included other objects needed." It is worth noting that many of the prompts asked students to compare the GenAI application to various works, so we had a high number of "compare" codes.

Conclusion

This study aims to shed light on the composition students' perceptions and interactions with GenAI tools integrated into writing instruction. While the small sample size limits generalizability, structured reflection revealed nuanced student attitudes about the quality, process, ethics, and efficiency of AI-assisted writing (Table 5.1).

Table 5.1. Summary of Codes

Major Code Category	Code Definition	Sub-Categories
Assessing based on **Process**	Students' claims, or predictions, of the usefulness of general and sometimes specific GenAI applications based on how they would help with specific parts of the writing process.	read, grammar, summarize, analyze, outline, brainstorm, find source
Assessing based on **Efficiency**	Students' assessment of how easy the GenAI application to use and/or how quick it functions.	easy (y/n), speed (y/n)
Assessing based on **Ethics**	Students' reflections on their understanding of the ethics and/or allowance of using the GenAI application.	confident, unsure, stores
Assessing based on **Quality**	Students' assessment of how well the GenAI application performed and the quality of the resulting data.	prompt (y/n), not enough, more (p/n), form (y/n), compare

Student evaluations of AI tools involved several key dimensions. First, in terms of the writing process, students assessed how well these tools supported specific stages, including reading, grammar, summarizing, analyzing, outlining, brainstorming, and source finding, often extending their use beyond what was explicitly assigned. Moreover, efficiency emerged as a significant factor in students' evaluations. While some students appreciated the intuitive and time-saving aspects of the tools, others found the need for re-prompting to be cumbersome and inefficient. Ethical considerations also played a crucial role, with students expressing varied levels of confidence and uncertainty about issues related to academic integrity, plagiarism, and the responsible use of GenAI. Lastly, the quality of the AI output was a critical focus, with students assessing its effectiveness based on prompt responsiveness, detail, adherence to genre conventions, and comparative evaluations against their own writing or other tools.

The results of this study generally align with scholarship about the various factors that impact adoption: personal traits, prior experiences, the nature of the innovation, perceptions of usefulness, ease of use, emotional responses, and contextual influences that evolve over time (Adler & Clark, 1991; Hall, 1979; Hord et al., 1987; Rogers, 1995; Zou & Huang, 2023). More importantly for writing instructors, these findings underscore the importance of providing intentional and reflective opportunities for students to navigate the evolving landscape of GenAI in academic writing. Kostopolus (2025) addresses the ethical dilemmas linked to integrating GenAI in multimodal composition, particularly those related to authorship, honesty, and intellectual property.

She argues that educators can effectively incorporate GenAI into multimodal composition by encouraging students to treat it as a tool that supports, rather than replaces, their writing, while remaining mindful of ethical citation practices. Assigning reflections helps students to understand where, how, and why they are using GenAI as a tool. When provided the opportunity to reflect, students can demonstrate they are critically nuanced users of writing technologies, especially GenAI applications.

References

Adler, Paul S., & Clark, Kim B. (1991). Behind the learning curve: a sketch of the learning process. *Management Science, 37*, pp. 267–281. https://doi.org/10.1287/mnsc.37.3.267

Al-Mamary, Yaser. Hasan., & Abubakar, Aliyu. Alhaji. (2025). Empowering ChatGPT adoption in higher education: A comprehensive analysis of university students' intention to adopt artificial intelligence using self-determination and technology-to-performance chain theories. *The Internet and Higher Education*, *66*, 101015. https://doi.org/10.1016/j.iheduc.2025.101015

Barron, Anne E., Kemker, Kate, Harmes, Christine, & Kalaydijian, Kimberly. (2003). Large-scale research study on technology in K-12 school: Technology integration as it relates to national technology standards. *Journal of Research on Technology in Education, 35*, pp. 489–508. https://doi.org/10.1080/15391523.2003.10782398

Bensalem, Elias, Harizi, Radhia, & Boujlida, Amel. (2024). Exploring undergraduate students' usage and perceptions of AI writing tools. *Global Journal of Foreign Language Teaching*, *14*(2), pp. 53-65. https://doi.org/10.18844/gjflt.v14i2.9344

Črček, Nikola, & Patekar, Jakob. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education*, *9*(4 (36), pp. 128-138. https://doi.org/10.17323/jle.2023.17379

Cummings, Robert E., Monroe, Stephen M., & Watkins, Marc. (2024). Generative AI in first-year writing: An early analysis of affordances, limitations, and a framework for the future. *Computers and Composition*, *71*, 102827. https://doi.org/10.1016/j.compcom.2024.102827

Gallagher, John. R., & Wagner, Kyle. (2024). Comparing student and writing instructor perceptions of academic dishonesty when collaborators are artificial intelligence or human. *Journal of Business and Technical Communication*, *38*(3), pp. 266-288. https://doi.org/10.1177/10506519241239937

Hall, Gene E. (1979). The concerns-based approach to facilitating change. *Educational Horizons, 57*, pp. 202–208. https://www.jstor.org/stable/42924345

Hord, Shirley M., Rutherford, William L., Huling-Austin, Leslie, & Hall, Gene. E. (1987). *Taking charge of change.* Association for Supervision and Curriculum Development.

Kostopolus, Emma. (2025). Student use of generative AI as a composing process supplement: Concerns for intellectual property and academic honesty. *Computers and Composition*, *75*, 102894. https://doi.org/10.1016/j.compcom.2024.102894

Maya, Dario. (2024, Spring). *AI comes for education: Can it make us better? Kairos Magazine*. Rutgers University School of Communication & Information. https://kairosmagazine.rutgers.edu/spring2024/ai-comes-for-education

Rogers, Everett, M. (1995). *Diffusion of innovations (4th ed.)*. Free Press.

Sergeeva, Olga. V., Zheltukhina, Marina R., Shoustikova, Tatyana, Tukhvatullina, Leysan R., Dobrokhotov, Denis A., & Kondrashev, Sergey V. (2025). Understanding higher education students' adoption of generative AI technologies: An empirical investigation using UTAUT2. *Contemporary Educational Technology*, *17*(2), ep571. https://doi.org/10.30935/cedtech/16039

Strzelecki, Artur. (2024). Students' acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology. *Innovative higher education*, *49*(2), pp. 223-245. https://link.springer.com/article/10.1007/s10755-023-09686-1

Zou, Min, & Huang, Liang. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology, 14*, 1259531. https://doi.org/10.3389/fpsyg.2023.1259531