Journal of Academic Writing
Vol. 15 No 2 WINTER 2025, pages 1-12
http://dx.doi.org/10.18552/joaw.v15i2.1346

Student Evaluative Judgements of Writing and Artificial Intelligence: The Disconnect between Structural and Conceptual Knowledge

Christopher Eaton
University of Toronto, Canada

Kaitlyn Harris
University of Toronto, Canada

Erin Vearncombe
University of Toronto, Canada

Abstract

This paper reports on how undergraduate students evaluated writing outputs created with and without generative artificial intelligence (AI). The paper focuses specifically on two aspects of writing and AI: how prior writing knowledge influenced students' thinking about AI tools, and how the writing skills to which they were exposed in the writing classroom helped them work with AI-generated materials. This research builds upon Bearman et al.'s (2024) work on evaluative judgement as a pedagogical tool to support learners as they work with AI-mediated texts. The paper uses this lens to identify challenges that learners have in applying writing knowledge to AI-mediated situations and to devise pedagogical means to support student learning in these contexts. We found that, while students could typically evaluate structural components of writing, they struggled to evaluate conceptual ideas both for AI and human generated texts. The findings speak more generally to the need for students to develop their evaluative abilities, as well as ways that AI may reveal and amplify existing challenges that learners have with evaluating the quality of writing, engaging with source materials, and applying genre knowledge to create meaning.

Introduction

The emergence of generative artificial intelligence (AI) is changing how students approach writing activities. Early AI discussions have focused on how learners evaluate AI outputs. In response, we drew upon Bearman et al.'s (2024) framework of evaluative judgement to analyze students' abilities to assess written outputs. We focused on two aspects of evaluation in particular: how prior writing knowledge influences students' thinking about AI, and how prior writing skills helped them work with written outputs.

This paper reports on results of research with 35 undergraduate students at a large Canadian university (September 2023-January 2024). We used constructivist grounded theory (CGT) as a methodological framework that could be adapted to suit the rapidly changing realities of AI.

Our research was guided by three primary questions:

1. How do former first-year writing learners think with and about AI tools as part of their writing activities and processes?

2. What writing skills do learners need to effectively integrate AI tools?

3. What writing skills can help learners make strong evaluative judgements in AI-mediated contexts?

The first two questions targeted students' prior writing knowledge and their dispositions towards AI. These questions established a baseline for how students' writing knowledge informed their AI use. Question 3 was formed after preliminary analysis highlighted the importance of evaluative judgements on students' abilities to use AI effectively.

Our findings suggest that there is a divide between how students evaluate the structure of writing and how they evaluate the conceptual frameworks underpinning written outputs. This divide was true whether the text they evaluated was generated by AI or a human. The findings indicate a need for learners to develop their evaluative abilities to successfully use AI to support their writing processes. The findings also suggest that evaluative challenges are not new. AI's influence amplifies existing challenges that learners have in evaluating the quality of writing, engaging with source materials, and applying genre knowledge to create meaning.

Literature Review

AI's rhetorical and genre implications have been a significant scholarly focus. Fyfe (2022), Morrison (2023), and Lingard (2023), for example, provided accounts of how writers could work with AI tools to refine their writing, noting that writers must ensure that outputs match rhetorical context. Jensen and Jensen's (2025) findings showed that students may benefit from using AI to find secondary sources, warning about a gap between how students use the technology (successfully) and the negative dispositions of their professors and librarians towards AI. Omizo and Hart-Davidson (2024) also underscored the need for genre knowledge to create and evaluate AI output. These authors tested the genre affordances of AI tools, usefully delineating how writers' prior knowledge of genre and writing could help them use AI. Moreover, Pigg (2024) examined the genre and rhetorical implications of using AI, emphasizing that 'the most effective writers working with generative AI content will need to dwell in the realm of evaluation and revision, drawing on both writing and content expertise to engage and critique output' (p. 8). This requirement for writing and content expertise places genre considerations at the forefront of the processes that writers may employ to leverage AI tools in their writing.

Evaluative Judgement and AI

Evaluation, or a writer's ability to assess quality (see Tai et al., 2018), has been emphasized in recent scholarship because of the iterative nature of AI use (e.g., Graham, 2023; Pigg, 2024) and AI's ability to generate text independent of human thinking (Omizo & Hart-Davidson, 2024). For learners who use AI in their writing processes, this ability to evaluate the quality of AI output is an important skill that they adapt for different writing contexts. This is where a pedagogical focus on evaluative judgement can help.

Evaluative judgment is especially useful because it could help learners navigate the uncertainty that is inherent in AI outputs, or the 'black box.' Bearman and Ajjawi (2023) argued that:

> Rather than explaining or unpacking 'black boxes'—things that we do not understand and cannot explain—our graduates must learn to work and make judgements within complex organisational and social structures, where they need to synthesize information in a sophisticated way as part of their professional practices. (p. 1162)

Learners must become arbiters of quality who verify, nuance, and develop AI outputs. Bearman et al. (2024) explained that 'good evaluative judgement is underpinned by a contextualised understanding of quality' (p. 7) and, to develop an understanding of quality, learners must 'appraise many different examples' (p. 7) within disciplinary contexts. They also note that learners do not always have the skills required to make good appraisals. These appraisals require genre knowledge: for good evaluative judgement to occur, a writer's claims must be defensible within the standards of the writer's context.

Methodology

Study Context

The research team included Christopher Eaton (Chris), Kaitlyn Harris, and Erin Vearncombe. Chris is an assistant professor, teaching stream and principal investigator (PI) on the research project. Kaitlyn was a graduate student and served as the research assistant for the project. Erin, also an assistant professor, teaching stream, joined the project after data collection to support the analysis. Data analysis and the writing of the manuscript were collaborative.

All 35 participants had successfully completed the university's first-year undergraduate writing course, which uses a 'writing about writing' approach (Downs & Wardle, 2007) that makes writing research central to assignment design and pedagogical framing. The framework emphasizes that learners should understand themselves as writers in their coursework, opening pathways for students to engage with writing metacognitively and develop transferable skills. Although the framework is popular in the United States, our course is the first of its kind in a Canadian context.

Our first-year course is a transitional, required course for most students across disciplines. Many students enroll in their first year of study, though classes often include students from upper years who take the course to fulfill degree requirements. Over 100 course sections are offered annually, each capped at 25 students. The sections draw from a syllabus with common assignments and core readings. The course begins with a reflective writing assignment, then proceeds to a discourse community analysis (expository), which builds into a genre analysis (argumentative). This genre analysis requires that learners conduct research on genre, rhetoric, and writing to analyze written communication within their discourse communities. It intends to provide learners with a foundational knowledge of genre that they can then adapt to future writing contexts. The third assignment's emphasis on genre knowledge made it an ideal model for the drafts participants engaged with in this project (see below) because of the importance of genre knowledge to evaluative judgements.

Recruitment began after receiving ethical approval in August 2023. Two calls for participants were made on our university's learning management system, one in August 2023 and one in December 2023. Thirty-three participants participated in all parts of the study, while two participants completed the interview only. Although it was not part of the inclusion/exclusion criteria, learners came from various disciplines (anthropology, biology, computer science, sociology, political science, chemistry, etc.). They also achieved a diverse range of grades in the course (from a C through to an A), though there were more participants from the B and A ranges. The time of recruitment meant that most participants completed the first-year course 6-12 months prior to participating.

Constructivist Grounded Theory

We used Constructivist Grounded Theory (CGT) as a methodological lens. CGT employs a series of flexible strategies that allow researchers to adjust to emerging data (Charmaz, 2014). Interviewing is often a cornerstone of this methodology because interviews capture participants' narratives of experience (Rieger, 2019). CGT does not have distinct stages for data collection and analysis. As Holton (2010) explains, data analysis and data collection inform each other. Data collected in the first stage of a project is analyzed, and that analysis shapes the next round of data collection. This process can repeat multiple times as the analysis narrows. The inductive approach that CGT espouses helped us adapt to emergent patterns in AI and participant experiences.

Methods

Semi-Structured Interview

The first phase included a one-hour semi-structured interview with each participant. Semi-structured interviews offer space to co-construct meaning as participants and interviewers can respond to emergent conversation threads (Charmaz, 2014). The interviews targeted the first two research questions, offering a baseline opportunity to understand how learners were thinking with and about AI tools as part of their writing processes. They also traced participants' prior knowledge about writing processes, writing skills, and how/if they saw this prior knowledge connecting to AI-mediated writing contexts.

Focus Groups

The interviews were followed by a two-hour focus group designed to see participants' evaluative judgements in action. There were seven focus groups total (4-6 participants each). Building on Bearman et al.'s (2024) emphasis on the appraisal of different examples of texts to facilitate evaluative judgements, each focus group was framed around a draft of a genre analysis assignment similar to the drafts participants developed in the course. Participants therefore completed tasks that asked them to draw on prior experience and competency.

Participants were given one of two possible drafts: one generated mostly by ChatGPT 3.5 with some minor syntactical and structural interventions from Chris, or one generated by Chris without AI. The content of the papers was similar: both were about AI's influence on writing processes. The papers also shared similar conceptual and structural challenges. Structurally, the drafts moved between topics quickly with many missing links between ideas, both in terms of metadiscourse and analysis. Conceptually, the papers had two major central flaws. The first was that the genres being analyzed and the arguments built around them were not commensurate, and the sources were incompatible and not part of the same scholarly conversations. The second flaw was that source engagements were cursory. The drafts drew on writing research but did not engage with the concepts that underpinned that research. Rather, the drafts would quote or paraphrase that which was most convenient for the argument. These issues are not uncommon for learners in this course when they submit a first draft. Maintaining the same cohesion between the drafts allowed us to examine participants' evaluative judgements through a similar lens regardless of what draft they received.

During the first hour, participants were asked to provide peer comments on a word document. Participants could converse as they developed comments, and each group was invited to discuss the drafts half-way through the hour. In the second hour, participants were asked to develop a step-by-step revision plan for how they would revise the draft, detailing their revision process and the skills/concepts from class they would draw upon.

Think-Aloud Protocol

During the second hour, each participant participated in a five-minute think-aloud session. Think-aloud protocols are brief recorded conversations that capture participant thinking as they developed their revision plan. In these five-minute sessions, participants were asked to describe their revision plan decisions expand upon the rationale for the importance of those decisions. This offered a glimpse into their decision making and offered complementary commentary that could add a dimension to our document analysis.

Think-aloud protocols also offer valuable methods to capture participant thinking in the moment and mitigate potential power imbalances (Rowell, 2022). There was an inevitable power dynamic between participants (undergraduates) and researchers (faculty). The letter of information and consent, as well as previous rapport helped mitigate these power imbalances. Rowell (2022) emphasizes that think-aloud protocols offer participants a chance to provide input on how their responses are interpreted, offering them a measure of control.

Analysis with CGT

CGT analysis is ongoing and happens simultaneously with data collection. Analysis is not a discrete stage of the research (Holton, 2010). It allows constant comparison and theoretical sampling (Breckenridge, 2012). Analysis began with open coding to gauge the depth and

breadth of their datasets (Holton, 2010). We read the interview transcripts, focus group documents, and think-alouds line-by-line, developing as many codes as possible. Following Qureshi and Ünlü's (2020) process, we combined the codes into concepts, descriptive terms that group similar ideas, before comparing and organizing these concepts into more abstract categories that ultimately developed into themes. The analysis moved systematically from initial description to broader theoretical insights. For example, codes such as 'using sources,' 'developing ideas,' 'connecting ideas,' and 'analyzing source quotes' were grouped into a concept we termed 'writing as a representation of thinking,' which was then organized within the larger category of 'learner epistemologies.' This category contributed to our thematic finding that learners tended to focus more on structural dimensions of genres rather than the underlying conceptual frameworks, as we discuss in the Results section.

We completed these analytical steps independently to begin. Charmaz and Thornberg (2021) emphasized the need to treat codes and categories as 'provisional and open for revision or rejection in light of new data and further analysis' (p. 322). Keeping our initial analyses separate built in this provisional approach because our individual analyses could be compared with two others.

After this initial analysis, we met to compare results. This stage allowed us to ask questions of our analysis, which Charmaz and Thornberg (2021) advised was useful to combine and make connections between. To add a layer to group analysis, we appointed Kaitlyn as a contrarian who prompted other analytical possibilities.

Analysis for Evaluative Judgement

To understand the implications for evaluative judgement, we needed to add a layer to our analysis. Our CGT analysis revealed differences in how learners applied their skills to structural and conceptual dimensions of evaluative judgements. By structural knowledge, we refer to their knowledge of mechanical, organizational, and structural features that comprise a text. By conceptual knowledge, we focused primarily on how learners worked with, represented, and nuanced ideas in the text.

To organize this part of our analysis, we developed analytical subcategories along which the codes and concepts could be examined. We drew upon Driscoll and Cui's (2021) writing knowledge, writing skills, writing strategies approach to trace different dimensions of what student writers have learned. Their approach offered a way to notice whether our participants recognized or could name the skills they used in their evaluations. Tardy's (2009) overview of how learners acquire genre knowledge (through formal knowledge, rhetorical knowledge, process knowledge, and subject matter knowledge) was the second element that informed these subcategories. Because genre knowledge is an important aspect of evaluative judgements, these subcategories helped to construct Table 1, which guided the final stages of analysis.

Conceptual awareness is a necessary component of evaluative judgement in academic contexts. While evaluative judgement refers to assessing quality relative to context (Bearman et al., 2024; Tai et al., 2018), making such judgements in disciplinary writing requires an understanding of how ideas connect, cohere, and function within scholarly conversations (see the 'Conceptual' category in Table 1, below). Without this conceptual awareness, learners identify surface-level problems but cannot assess whether ideas are valid or well-reasoned. Conceptual awareness develops through engagement with disciplinary content and genre conventions; these elements cannot be separated (Driscoll et al., 2020).

Table 1: Breakdown of Analytical Sub-Components

Broad Categories from CGT Analysis	Sub-Component	Description
Structural	Paragraph Structure	Knowledge of elements like topic sentences, examples, placing and building analysis in the paragraph structure
	Integrating Source Material	The structure that source engagements assume and how they transition to other parts of the discussion
	Developing Introductions	Creating space to articulate the core problem/the questions being asked, to contextualize the paper and establish an analytical thread
	Metadiscourse	Sentence-level features such as hedging, they say/I say features
Conceptual	Understanding of Key Concepts	The accuracy, depth, and breadth of understanding demonstrated around important writing concepts/theories/ideas (e.g., genre, rhetoric)
	Analyzing Source Material	Focus on the quality, accuracy, and depth of the way a paper joins the scholarly conversation with specific sources
	Argument	Focus on the consistency of a text's arguments/claims and how they align with the wider conversation
	Contextual Understanding	How well they learn, know, and account for the contextual demands of a genre or writing task

We attributed attribute specific codes or concepts to a sub-component, and we noted the participants to which they corresponded. This analytical stage was fluid, with some codes fitting into multiple categories. This flexibility allowed us to determine which sub-components arose most frequently and contributed the most to the broader categories. Our goal was to notice patterns between the sub-components and the larger categories. The results and discussion that follow are guided by the most prominent patterns that emerged.

Findings

Implicit Standards for Evaluative Judgement: Insights from the Interviews

The semi-structured interviews examined how undergraduate learners situated AI within their writing processes. Our analysis revealed that evaluative judgement mediates interactions between student writers and LLMs, but students seemed unaware of how and why they make specific evaluations.

Students reported using AI as a 'tool' or 'assistant' to augment their thinking and increase their writing efficiency. Common functionalities included summarizing sources, generating ideas and counterarguments, producing outlines, rephrasing sentences, and checking grammar. Participants employed these AI affordances recursively: inputting a prompt, receiving outputs, evaluating their validity, and using resulting insights to refine subsequent prompts. Many inputted the same prompt multiple times to receive varied outputs they could adopt, adapt, or synthesize. These findings align with the AI processes outlined in Graham (2023), mirroring the 're-writing and revision' (p. 166) cycles that characterize writing processes more generally.

The success of these practices depends on the writer's ability to make evaluative judgements about output quality relative to genre expectations, contextual demands and disciplinary standards (Bearman et al., 2024): understandings developed through genre-based knowledge (Omizo & Hart-Davidson, 2024; Pigg, 2024). To remain epistemic agents, students must develop a way to discern the structural, conceptual, and contextual appropriateness of texts mediated by AI.

When asked how they evaluated the outputs they received from AI, most participants struggled to articulate any quality standards guiding their decision-making. Over 25% could not name a single writing-related concept or competency that applied to their work with AI. Those who engaged with the evaluative process were equally unable to offer specific strategies, instead reporting reliance on 'intuition' developed through repeated exposure to texts of a similar genre. This insight corresponds to Bearman et al.'s (2024) observation that a writer's criteria for evaluative judgements are often implicitly and holistically understood.

In human-mediated writing contexts, tacit understanding of genre considerations and quality standards may be relatively functional. However, managing AI's potential for inaccuracy, omission, and hallucination requires domain-specific knowledge (Bearman & Ajjawi, 2023). After conducting and analyzing the interviews, we needed additional information to gauge how writers went about making evaluative judgements with human- versus AI-generated texts.

The Structural/Conceptual Divide: Trends from the Focus Groups and Think-Aloud Protocols

The focus groups and think-aloud protocols allowed us to consider the actual criteria by which participants evaluated papers from AI and human authors. Our primary finding was that the students lacked conceptual-level engagement with the texts under revision. While they understood the structural need for certain rhetorical strategies, participants rarely considered the coherence of these features within the papers. Consequently, the students did not meaningfully adapt their approaches to address the conceptual shortcomings of AI-generated texts.

Structural Awareness—Organization and Form

The first category to emerge was structural awareness: how textual features (e.g. topic sentences, quotations, transitions) should be employed within academic essays. Evaluating structural validity requires intersecting formal and rhetorical knowledges (see Tardy, 2009). Experienced students may hold such understandings intuitively, using explicit frameworks only to retroactively explain their evaluative process (Tai et al., 2018).

Participants were generally successful at applying structural knowledge to the revision task, regardless of perceived authorship (AI or human). Participants consistently identified organizational flaws, with suggestions concentrated around adhering to a central argument, crafting clear topic and concluding sentences, and maintaining a logical cohesion between paragraphs. Many also noted that the sentences following a quotation should connect the source back to the paper's primary claim. In expressing these ideas, participants explicitly called upon concepts and terminology learned in the first-year writing course (e.g. topic-gap-argument, beads-on-a-string, they say/I say). Although unable to articulate any evaluative criteria in interviews, all participants operationalized at least one such quality standard in practice. The revision task helped 'intuitive' understandings to become overt, revealing strong structural awareness with both human- and AI-generated texts.

Conceptual Awareness—Contextual Understanding

Our second analytical category was conceptual awareness: how ideas interact to create generative meaning beyond form and organization (Perkins, 1993). Students need subject matter and process knowledge to understand how these ideas are produced, positioned, and communicated in the target genre (Tardy, 2009). While participants provided thorough structural revisions, commentary on conceptual coherence was often missing altogether.

The first conceptual skill that students overlooked was adapting to rhetorical context. Tai et al. (2018) emphasize that quality standards are always defined "in relation to a problem space" (p.

472). Participants widely understood how genre and context inform evaluations, frequently mentioning in interviews the need to compensate for the isolated nature of AI-generated text. Participant 17 commented that a human writer must bring 'a strong understanding of [their] assignment and the concepts [they] need'—with participant 13 adding that this awareness is necessary because 'humans have better understandings of context than an AI would.' Students largely recognized that serving as arbiters of quality (Bearman et al., 2024, Tai et al., 2018) necessitated attending to externally-defined expectations.

However, focus groups participants did not implement corresponding strategies for either the AI or human paper conditions. Each group received print copies of the assignment prompt and rubric, but most never consulted them. As a result, only five out of the 33 focus group participants identified that the papers failed to fulfill a primary assignment expectation: situating genre within a discourse community. Participants' disregard for assignment instructions implies that they held existing notions about successful papers without recognizing that their understandings were incomplete or that AI's potential influence may have shifted the composition context. Although revisions standardized the papers' structure and organization according to the essay format, they did not help to achieve rhetorical purpose.

Conceptual Awareness—Analyzing Source Material

Participants' second conceptual struggle was analyzing source material. This struggle does not refer to the logistics of inserting and unpacking source quotations, which our participants handled capably as part of their structural evaluations. Instead, this sub-component involves judging whether source use contributes to credible, coherent arguments within disciplinary space (Bearman et al., 2024).

While participants collectively produced 36 comments about sources, most revisions focused on structurally integrating evidence with analysis. Some students expressed an intention to check the sources, but most comments prioritized checking a source in an individual sentence or idea in a paper. Participants mostly did not consider the relationship between sources in the scholarly conversation, and they did not indicate that they would check whether the sources were a good fit for the paper's central claims. Only one of 33 participants identified another primary flaw: secondary sources were fundamentally incompatible with each other. In both the AI and human paper groups, students' skepticism applied in only a limited capacity.

This lack of conceptual analysis reveals a challenge with how undergraduate writers manage uncertainty in 'black box' evaluative situations. They used genre awareness to improve text structure (whether human or AI-authored) and recognized the need to verify claims outside the scope of their knowledge. With these responsibilities completed, little reflection addressed the bigger picture: whether claims were individually accurate *and* collectively coherent. Quality standards were structural rather than conceptual; discussion around meaning-making or knowledge construction was conspicuously absent. When faced with opaque logical processes, students had difficulty applying evaluative judgement to conceptual revisions.

Discussion

Challenges with Conceptual Judgements

The data indicated a divide between students' structural and conceptual evaluative abilities. They applied elements related to structure, organization, and mechanics but struggled with conceptual-level text engagement. In focus groups, participants focused on what looked consistent with their experience writing genre analyses, revising to build structure (like refined transitions between sources and analysis) around the draft's existing foundation. That participants ignored the rubric and assignment details implied that they overestimated their prior understanding. As a result, they worked with the draft's established ideas rather than scrutinizing the ideas themselves. They could target individual structures for more consistent discussion (like refining the paper's central thread and building in a thesis) but did not connect those revisions to rhetorical demands.

These assessment challenges reveal how participants took up genre knowledge. Tardy (2009) identified four dimensions of genre knowledge: formal (genre conventions, components), rhetorical (genre strategies, functions), process (genre production, distribution, consumption), and subject matter (disciplinary content). These dimensions enable learners to understand genre features and functions while recognizing how knowledge is contextually and disciplinarily produced. Participants demonstrated some genre knowledge and demonstrated a grasp on rhetorical dimensions comprising the genre. However, they struggled to apply subject matter knowledge and lacked crucial process knowledge. Rather than appraising the logic behind the subject matter and source engagements (the two central flaws in the drafts), they immediately zoomed into the syntactical/organizational dimensions, assuming the draft's ideas were valid.

Evaluative Judgement

The draft participants received required them to develop a strong conceptual understanding of the draft before focusing on the final structure. Most participants, however, seemed to lack the evaluative skills to do this well. That only five of 33 focus group participants realized that the draft did not situate the genre within the discourse community is one example of this challenge. This lack of evaluative judgement is also evident in that only one participant realized that the core source texts were incompatible. As Bearman et al. (2024) noted, learners do not always have the skills necessary to make good appraisals. It is arguable that these learners had the requisite genre knowledge to make good appraisals for the focus group drafts—at least in theory. Their prior exposure to genre research could have given them the skills necessary for this task. However, what they did not have was the prior knowledge about AI's influence on writing contexts, and our first-year class provided few experiences assessing outputs. Peer feedback activities in class offered chances to support others and help them construct around the ideas they had established, but these activities did not seem to prepare learners to scrutinize the concepts that underpin a discussion.

It is noteworthy that learners with human generated drafts equally had a difficult time moving beyond the surface and engaging conceptually. While AI may have been a factor in these conceptual issues, the AI impetus for the study revealed a challenge that learners may have with evaluative judgements more generally. This challenge is even more prevalent with the influence of AI tools because of the nature of human-AI interactions. Humans will constantly be faced with situations where ideas cannot be traced back to their source. As Bearman et al. (2024) explained, good evaluative judgement requires that a person/writer's claims are defensible within their rhetorical situation. For writing pedagogies to account for AI's influence, drawing attention to the validity or commensurability of ideas will be useful. It will be important for learners to recognize how, epistemically, AI-mediated text design places different demands on writers. They must also recognize how their evaluation of texts must adjust to account for the possibility of AI's influence. This involves not simply working with an idea, as our participants did with the drafts. Rather, it involves recognizing the deeper conceptual constructions that underpin a text and how meaning is made within a given context. From there, learners must know how to evaluate those constructions and know when to adjust their design process to suit the context.

One other dimension could be useful for learners to develop: the ability to scrutinize three levels of certainty to promote evaluative judgement when evaluating potentially AI-mediated texts. These levels include: (1) what learners can confirm as true based on prior knowledge; (2) what they do not know is true but can confirm/verify based on information available; (3) what they cannot confirm is true. The third level accounts for the 'black box,' where some elements may never be known, and learners must learn how to navigate this uncertainty. Much of this work requires that learners recognize the limits of their own knowledges on Tardy's (2009) notion of formal, rhetorical, process, and disciplinary planes. Undergraduate learners are better positioned to apply prior writing knowledge and skills when they can recognize when their prior knowledge and skills are sufficient or insufficient for the current task.

Chris has since applied these ideas to his pedagogy. In his first-year classes and in an upper-year seminar on AI, he invites students to grapple with these three levels of certainty through a design activity: learners annotate a draft paper, indicating what information they know is true, what they can confirm is true, and what they cannot confirm. In the upper year course, the draft

is generated with support from AI because of its alignment with course learning outcomes. In the first-year course, learners use their own draft of the genre analysis before submitting; this positions them to strengthen the ideas that are central to their paper. After their annotations, participants are asked to describe why they made specific annotations. This reflection is designed to encourage learners to confront the limits of their own knowledge and, potentially, challenge their evaluative judgements.

Conclusion

In writing contexts where machines produce content independent of humans, novice undergraduate writers' limited capacity for strong evaluative judgements creates a barrier to their writing development. Our findings indicate that successful content evaluation, whether AI-generated or not, requires learners to critically engage with the epistemological processes of their rhetorical situations. In other words, they must possess genre knowledge, including subject matter knowledge: content matters. While the importance of content knowledge in writing is well-established, this study adds a fundamental connection: evaluative judgement in AI-mediated contexts depends heavily on writers' genre knowledge and their ability to assess whether content meets disciplinary and contextual expectations. An important next step will be to better understand the levels of knowledge learners require before being in a position to make strong evaluative judgements. To reach this understanding, it will be important to consider both disciplinary knowledge and AI knowledge.

Going forward, it is worth asking whether the challenges undergraduates face evaluating written outputs are created by AI, exacerbated by AI, or represent longstanding issues in writing development that AI renders more visible. Many writing classrooms build on the premise that writing is an acquired technology of communication and an imperfect means of cognitive translation and inscription requiring time, practice, and tools for effective development. AI is emerging as another tool within this ever-evolving technological apparatus. As such, generative AI may be another iteration of longstanding challenges learners face in developing and adapting writing skills to new contexts. Rather than creating entirely new pedagogical challenges, AI may illuminate persistent difficulties novice academic writers face. Understanding this distinction is crucial for developing writing pedagogies that strengthen students' evaluative judgement and genre knowledge, foundational skills that will serve them well when they encounter AI tools in academic settings and beyond.

References

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology,* 54(5), 1160-1173. https://doi.org/10.1111/bjet.13337

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment and Evaluation in Higher Education*, 1-13. https://doi.org/10.1080/02602938.2024.2335321

Breckenridge, J.P., Jones, D., Elliott, I., Nicol, M. (2012). Choosing a methodological path: Reflections on the constructivist turn. *Grounded theory review: An international journal, 11*(1), 64-71. https://groundedtheoryreview.com/2012/06/01/choosing-a-methodological-path-reflections-on-the-constructivist-turn/

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Sage.

Charmaz, K., & Thornberg, R. (2021). The pursuit of quality in grounded theory. *Qualitative Research in Psychology, 18*(3), 305-327. https://doi.org/10.1080/14780887.2020.1780357

Downs, D., & Wardle, E. (2007). Teaching about writing, righting misconceptions: (Re)envisioning "first-year composition" as "introduction to writing studies". *College Composition and Communication, 58*(4), 552-584. https://doi.org/10.58680/ccc20075923

Driscoll, D. L., Paszek, J., Gorzelsky, G., Hayes, C. L., & Jones, E. (2020). Genre knowledge and writing development: Results from the Writing Transfer Project. *Written Communication*, *37*(1), 69–103. https://doi.org/10.1177/0741088319882313

Driscoll, D.L., & Cui, W. (2021). Visible and invisible transfer: A longitudinal investigation of learning to write and transfer across five years. *College Composition and Communication, 73*(2), 229-260. https://doi.org/10.2307/ccc202131712

Fyfe, P. (2022). How to cheat on your final paper: Assigning AI for student writing. *AI & Society, 38*, 1395-1405. https://doi.org/10.1007/s00146-022-01397-z

Graham, S.S. (2023). Post-process but not post-writing: Large language models and a future for composition pedagogy. *Composition Studies, 51*(1), 162-168. https://compstudiesjournal.com/wp-content/uploads/2023/06/graham.pdf

Holton, J.A. (2010). The coding process and its challenges. *Grounded Theory Review: An International Journal, 9*(1). https://groundedtheoryreview.com/2010/04/02/the-coding-process-and-its-challenges/

Jensen, T. W., & Jensen, S. W. (2025). AI literacy in the context of working with sources: Pitfalls and possibilities of generative AI models in academic writing. *Journal of Academic Writing*, 15(S2), 1-12. https://doi.org/10.18552/joaw.v15iS2.12241

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education, 12*(1), 261–270. https://doi.org/10.5334/pme.1072

Morrison, A. (2023). Meta-writing: AI and writing. *Composition Studies, 51*(1), 155-161. https://compstudiesjournal.com/wp-content/uploads/2023/06/morrison.pdf

Omizo, R., & Hart-Davidson, B. (2024). Is genre enough? A theory of genre signaling as generative AI rhetoric. *Rhetoric Society Quarterly, 54*(3), 272-285. https://doi.org/10.1080/02773945.2024.2343615

Perkins, D. (1993). Teaching for understanding. *Professional Journal of the American Federation of Teachers, 17*(3), 28-35. https://www.aft.org/sites/default/files/media/documents/2023/17.3_fall_1993_reduced_size.pdf

Pigg, S. (2024). Research writing with ChatGPT: A descriptive embodied practice framework. *Computers and Composition, 71*. https://doi.org/10.1016/j.compcom.2024.102830

Qureshi, H.A., & Ünlü, Z. (2020). Beyond the paradigm conflicts: A four-step coding instrument for grounded theory. *International Journal of Qualitative Methods, 19*, 1-10. http://journals.sagepub.com/home/ijq

Rowell, C. (2022). Multimodal methods for mapping multimodal composing processes. In E. Del Hierro & C. VanKooten (Eds.), *Methods and methodologies for research in digital writing and rhetoric: Centering positionality in computers and writing scholarship* (Vol. 2, pp. 185- 208). WAC Clearinghouse. https://wac.colostate.edu/docs/books/positionality/volume2.pdf

Tai, J., Ajjawi, R., Boud, D., Dawson, P., Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*, 467-481. https://doi.org/10.1007/s10734-017-0220-3

Tardy, C.M. (2009). *Building genre knowledge*. Parlor Press.