



Making AI annoying on purpose: when helpful tools don't always help[☆]

Lindsay Konradt^{a,b,1,*}, David N. Boote^{a,1}, Michelle Taub^{a,1}

^a Department of Learning Sciences & Educational Research, University of Central Florida, United States

^b Orange County Public Schools, Orlando, Florida 32801, United States

ARTICLE INFO

Keywords:

Generative artificial intelligence
Argumentative writing
Self-regulated learning
Affordance theory
Writing pedagogy
AI design
Prompt literacy

ABSTRACT

This design-based study examined how four high school students used a constrained AI writing system comprised of teacher-delimited AI chatbots designed to ask questions rather than generate text to build argumentation skills. In a nine-week unit, the TRACE model (Target, Refine, Assess, Cycle, Extend) oriented interactions through specialized chatbots that promoted productive friction by requiring students to defend and refine their thinking. Data included writing samples, AI interaction logs, weekly reflections, and interviews analyzed through thematic analysis.

Analysis showed substantial changes in argumentative reasoning as students evolved from passive AI consumers to strategic evaluators, actively pushing back on AI outputs. They developed "prompt literacy," crafting targeted requests and critically evaluating responses, while strengthening counterargument development and evidence integration.

These findings challenge recent criticism of AI supported writing instruction and claims that AI's generative potential expands invention and play. In contrast, this study demonstrates that constraint, not unrestricted generativity, was associated with increased self-regulation and sustained engagement with argumentative contexts. Strategic AI limitations made sustained cognitive engagement unavoidable, embedding repeated evaluation into everyday composing decisions through productive friction between tool assistance and intellectual autonomy.

1. Introduction

Generative AI has exposed longstanding gaps in writing instruction, particularly in argumentative complexity and evidence synthesis. These struggles predate AI; research showed students struggled with this type of writing. Studies in secondary and higher education found writers produce weak, formulaic counterarguments instead of substantive ones [1,2]. Similarly, evidence integration remains shallow, with students listing sources sequentially rather than synthesizing them [3].

Argumentative writing demands sustained self-regulation across interconnected processes: planning arguments, monitoring reasoning quality, and revising based on evaluation of logical gaps [4–6]. These regulatory demands challenge student writers, who struggle not only with content generation but with the metacognitive orchestration of reasoning. Skilled instructors preserve student ownership by creating spaces and supports that refine reasoning [7,8].

Generative AI creates a pedagogical paradox. While proponents emphasize its potential for invention and creativity, writing theorists warn that helpful technologies can eliminate the cognitive work essential for learning [9]. Unrestricted AI access promotes cognitive off-loading, undermining evaluative and reflective processes [10,11]. The challenge: designing AI interactions that preserve cognitive work while providing meaningful support. We addressed this tension by designing AI writing tools with strategic constraints. By limiting generative outputs and restricting AI to questioning, educational environments can preserve the cognitive work that drives learning while providing meaningful support. This constraint-as-affordance approach creates productive friction: deliberate impedance that requires sustained logical engagement instead of passive consumption of suggestions, maintaining cognitive load essential for deep learning while leveraging AI's capacity for targeted intervention. This study's primary objective is design: examining a constraint-based writing support system in a classroom

[☆] This manuscript is a substantial revision of research that was originally published in Lindsay Konradt's doctoral dissertation.

* Corresponding author at: Department of Learning Sciences & Educational Research, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816. E-mail address: li112090@ucf.edu (L. Konradt).

¹ All authors are affiliated with the Department of Learning Sciences & Educational Research, University of Central Florida. Lindsay Konradt is also a teacher with Orange County Public Schools, Florida

setting. We investigate how limiting AI output to questioning reshapes student interaction with feedback during extended argumentative writing. Changes in writing practices and self-regulatory behaviors are presented as design outcomes informing viability and consequences, not as claims about generalized developmental gains. This study addresses the research question: How does a constraint-based AI assisted system influence argumentative writing development across extended instruction? Instead of measuring preferences or short-term engagement [12], we examine a specific constraint: limiting AI output to questioning only. This targets three aspects of argumentative writing: crafting strong counterarguments, engaging in sustained revision practice, and developing awareness of how to evaluate and respond to feedback. Instruction occurred during a nine-week unit with two sections of AP Capstone; data collection and analysis focused on four advanced high school students.

2. Literature review

Revision often remains surface-level, bypassing meaningful restructuring [7], "knowledge-telling" rather than "knowledge-transforming" [13]. These difficulties stem from self-regulation challenges: students bypass the metacognitive reasoning and focus on procedural completion.

2.1. Digital writing, AI tools, and cognitive offloading

Concerns about technology's impact on writing are not new. For decades, scholars have argued that digital tools fundamentally reorganize writing and writers' thinking [9,14]. Each technological wave has spawned the same pedagogical conundrum: do these tools extend human thinking by automating low-level tasks, or do they offload the cognitive work needed to build expertise?

Generative AI intensifies this question. While the technology offers promise for personalizing instruction and fostering innovative pedagogy [15,16], writing tools may inadvertently automate the cognitive processes students need to develop [17]. This creates two challenges: cognitive offloading and automation bias.

Cognitive offloading occurs when a writer uses a tool to perform essential cognitive labor, outsourcing the thinking required for learning [18]. When AI generates counterarguments or restructures paragraphs, it can eliminate the "germane cognitive load," the productive mental effort that builds transferable argumentative strategies [19]. This is compounded by automation bias, an inherent trust in AI output leading to passive acceptance [20]. This bias persists even among experts [21] and can lead to decreased student agency and "metacognitive laziness" that hinders argumentative development [22,11]. When students accept AI suggestions uncritically, the tool ceases to be a support and becomes a substitute for learning.

2.2. A framework for instructional design: balancing agency and constraints

Gibson's affordance theory suggests that AI's educational value emerges from the relationship between learner capabilities and environmental constraints, not from tool features [23,24]. This means AI's educational impact depends on how learning contexts shape tool use. For argumentative writing, the goal becomes designing AI interactions that preserve cognitive processes of planning, evaluating, and revising [10].

Strategic constraints can enhance student agency by making certain cognitive processes unavoidable [25]. When AI tools refuse to complete students' thinking, they create resistance that requires sustained engagement with argumentative reasoning. This challenges the assumption that helpful tools should eliminate cognitive effort, suggesting that strategic limitations can foster the self-regulatory processes central to writing development.

Digital writing instruction has centered on agency and cognitive

work. Full automation threatens voice; the challenge is democratizing creativity while improving discernment, including rhetorical ethics [26]. Students must learn to revise while interrogating AI-generated content and asserting voice amid algorithm-created suggestions. In the right context, AI can shape invention and iteration while centering reflection and judgment [27,28].

2.3. The empirical gap

Empirical support for constraint-based AI design remains sparse but conceptually grounded. While Socratic-style prompting and reflection-focused tutoring [29] show promise, most studies do not isolate the effects of limited output design. AI tutors prompting reflection fostered more sophisticated revision strategies than correction alone [30], but constraint was not isolated. In a broader study on student agency and self-regulated learning, Darvishi et al. [31] found that hybrid models combining AI feedback with explicit SRL instruction were no more effective than AI assistance alone, raising questions about how these supports interact.

Research on AI writing tools emphasizes their capabilities over learning processes. Like earlier educational technology innovations [32], this emphasis leaves gaps in understanding of argumentative growth. The limitation matters because rhetorical competence develops through iterative practice; isolated encounters do not establish durable writing habits, self-regulation, or ethical reasoning in real academic contexts [33]. Research emphasis on user satisfaction over learning outcomes limits understanding of how constraint-based design influences counterargument reasoning, evidence evaluation, and rhetorical awareness [34]. Established writing pedagogy research prioritizes measurable skill development over user experience metrics [35], yet AI research has not followed suit.

3. Methods

Design-based research methodology [36] guided our investigation of how constrained AI tools shape argumentative development. The TRACE model (explained below) functioned as an instructional and analytic organizing framework to guide student interactions with AI tools; an observable set of practices enacted through prompts, revisions and reflections.

3.1. Context

The study was conducted during a nine-week argumentative writing unit in AP Capstone Seminar at a public high school in the Southeastern US. All 28 eleventh-grade students used the constrained AI tools as part of regular instruction. AP capstone policies prohibit individual feedback after the first semester, creating a pedagogical need for alternative feedback sources during extended writing projects.

3.2. Participants

Four student volunteers were selected through maximum variation sampling based on baseline writing ability and AI attitudes. Baseline writing ability was determined through an initial sample essay with curated sources. AI attitudes gauged from prior classroom discussions, captured differences in openness, skepticism, or reliance on technology, which may have influenced uptake and resistance.

3.3. Instructional design and implementation

3.3.1. Curriculum structure and trace implementation

The curriculum followed a three-phase structure with progressive AI limitations designed to create resistance while sustaining support. This progression aligned with affordance theory's principle that tools should make cognitive engagement unavoidable, [23], gradually increasing

metacognitive demands as students developed self-regulatory competence.

TRACE (Target, Refine, Assess, Cycle, Extend) functioned as an analytic organizing framework for this study, not as an instructional protocol taught to students. It addressed self-regulatory challenges central to argumentative writing [5,6] by providing a lens for categorizing observable behaviors: targeting specific writing needs, refining approaches based on feedback, assessing tool usefulness, cycling through revisions, and extending strategies to new contexts. While TRACE was not taught as a formal sequence or checklist, the instructional design—including tool constraints, prompt requirements, and reflection questions—was structured to make these self-regulatory behaviors more likely to occur and easier to observe.

In Weeks 1–3, students had unrestricted access to AI-supported literature search tools (Elicit and SciSpace) while becoming familiar with expectations for AI use and evaluation. Weeks 4–7 introduced constrained tools gradually: the devil's advocate bot for claim refinement, followed by rhetorical audience and solution limitation tools. Strict prompt limits were set, with the number of prompts allowed decreasing weekly. Lex.page functionality was restricted from text generation, and students were required to document and evaluate tool feedback. By Weeks 8–9, students faced tighter prompt limits and had to justify tool use in reflections. As tool access decreased, metacognitive demand increased, reinforcing TRACE aligned practice as a shared instructional structure for students' evaluative decisions during writing.

TRACE informed all tool interactions, though with varying levels of formality. Weekly formal reflections through MirrorTalk required explicit documentation of their process: how they targeted specific writing needs, refined their approach, assessed AI feedback, cycled through revisions, and extended insights to new contexts. TRACE labels were applied analytically to categorize observable self-regulatory behaviors in students' work, not to assess whether students consciously followed TRACE phases or learned TRACE as a strategy. Coding focused on whether actions reflected evaluation or regulation (e.g., rejecting feedback with justification, revising a prompt after assessing usefulness, adjusting strategy based on prior assessment), regardless of whether students referenced TRACE terminology. Students were not taught TRACE labels; we used TRACE to make sense of the self-regulatory patterns that emerged.

TRACE was designed for this study's AI-supported writing context, but it draws on self-regulated models like Self-Regulated Strategy Development (SRSD, [37]). Like SRSD, TRACE emphasizes recursive planning, evaluation, and revision to strengthen students' agency in managing the writing process. However, where SRSD provides validated protocols for teaching broad writing strategies, TRACE focuses on scaffolding students' interactions with AI tools. Unlike SRSD, TRACE did not include a full strategy instruction sequence; it emerged contextually to guide students' interaction with constrained AI outputs.

3.3.2. AI tool constraints

The decision to restrict AI responses to questioning was deliberate, grounded in the instructional context and documented patterns of AI misuse in student writing. In AP Capstone specifically, students often offload argumentative reasoning to generative AI under grading pressure (personal observation; [38]). Question-only design directly addresses this by making cognitive delegation structurally impossible while preserving access to targeted feedback.

Alternative constraint designs exist and merit investigation: AI that provides tentative answers then challenges students, or phased approaches where constraints relax as competence develops. We did not pursue these because even "tentative" answers can function as cognitive anchors students defer to under time pressure, and our nine-week timeline with AP grading restrictions limited opportunities for phased relaxation. Question-only served as a starting constraint appropriate for this population (verbal, college-bound students with access to teacher modeling) and this context (extended argumentative writing with high

accountability). Future research should examine how alternative constraint designs function across different learner populations and writing contexts.

Four custom AI bots (outline, editing assistant, audience perspective, and devil's advocate) were programmed to reject text generation requests (Fig. 1). These constraints were consistent across tools and remained active throughout the unit. This approach builds on research showing AI's effectiveness in debate tutoring through line of inquiry [39]. This constraint-as-affordance design preserved the cognitive work that drives learning while providing meaningful support, preventing the automation bias that bypasses planning, monitoring, and revision cycles [20]. The devil's advocate chatbot questioned students in three ways:

1. Counterpoint questions introduce alternative viewpoints or challenge assumptions within an argument;
2. Socratic questions elicit deeper reasoning, prompting additional explanation or justifications; and
3. Further investigation questions encouraged students to identify topics needing more research or clarification.

The exchange in Fig. 1 occurred later in the unit, after students developed more specific prompting strategies. Betty's request reflects growing confidence and understanding of tool use.

Each chatbot was designed to address a specific argumentative function. Devil's advocate bot prompted counter positions, challenged assumptions, and requested justification for claims. Audience perspective bot posed questions about stakeholder values, rhetorical impact, and audience-specific consequences. Outline bot questioned claim hierarchy, logical sequencing, and alignment. Editing assistant focused on clarity, coherence, and orientation with genre conventions, without proposing rewritten text. Each tool followed a fixed questioning pattern; the system did not adjust its behavior based on perceived student ability, prior performance, or past interactions. Adapting to tool use was instructional; the AI tools themselves did not adapt to individual students; instead, students were required to adapt their use of the tools by becoming more precise, selective, and evaluative in their requests, and providing more context for their specific topic and task.

The other tools, SciSpace and Lex.page, were similarly constrained. In addition to teacher-imposed limitations, inherent system restrictions, such as excessive congeniality and potentially misleading feedback, afforded opportunities for divergent and critical thinking when chatbots did not function as expected.

3.3.3. Instruction and ethical integration

The design was pedagogical as much as technical; constrained tools required explicit instruction, modeling, and guided practice. Instruction followed a gradual release model aligned with cognitive apprenticeship principles [40,41]: teacher modeling with live demonstrations of prompt crafting, guided class practice with generic topics, then individual use with feedback. This approach preserved the cognitive work of learning to interact strategically with AI while providing scaffolded support. Students learned to distinguish between imprecise and specific inputs by seeing how constrained bots responded differently based on clarity and intent, developing the "evaluative reasoning" that effective writing instruction requires [8].

Ethical boundaries were embedded into the instructional design, following principles of experiential ethics. Students signed an AI Use Agreement outlining expectations: no full-text generation, no sentence copying, and all tool use must be documented and explainable. The first author narrated decision-making during demonstrations, reinforcing boundaries between AI support and replacement. Automatic acceptance of outputs was not allowed. Students were expected to evaluate tool feedback against their goals, preventing processes that undermine argumentative development [22,42].

Weekly reflections asked students what the tool helped them improve, which suggestions they rejected, and when it felt like the tool



Welcome to Devil's Advocate! I'm here to challenge your arguments and help you strengthen your reasoning skills. Let's dive into some thought-provoking discussions!



I would like for you to help me find weaknesses in my counterargument and rebuttal section of my research paper on how FOMO does not affect teens. On the other hand some people may argue that FOMO does not have any effects on teenagers in high school. The fear of missing out can be caused by certain factors such as social media but this can also be a positive experience. FOMO can be a good thing, and it can motivate a person in specific situations. (Winick)



Great start, Let's analyze your counterargument and rebuttal.

1. **Counterpoint to Assumption:** You mention that FOMO can motivate people positively, but how do you address potential adverse effects? Are there studies or examples that show how this motivation might lead to unhealthy competition or anxiety?
2. **Evidence Clarity:** The citation (Winick)" needs content. What does Winick argue specifically? Consider clarifying their viewpoint and how it supports your stance. How robust is this evidence?
3. **Broader Perspective:** While you focus on positive aspects, could you explore broader implications? For instance, have other studies pointed out the nuances of FOMO that contradict your argument?

Fig. 1. Devil's advocate exchange (Betty).

was doing too much. These prompts reinforced the self-regulatory processes central to argumentative development [6], leading students to articulate personal boundaries about preserving their voice versus accepting AI assistance. The ethical framework operated through transparency and repeated decision-making consistent with ecological affordance theory [23]. Because tools could not generate complete text, every interaction required students to decide how much help to accept while keeping ownership.

3.4. Data collection

Data collection occurred across the nine-week unit through multiple sources designed to track argumentative writing development and tool adaptation:

Writing Samples: Students completed baseline argumentative essays (1.5–2.5 pages) from teacher-provided sources, scored using the AP Seminar rubric. Throughout the unit, students produced ten iterative drafts on a self-selected topic, submitted through Canvas LMS: five status checks and their revisions, each building on previous sections while revising earlier content. This generated substantial writing data as students developed and refined seven drafts and revisions of major paper sections across multiple revision cycles. The final draft was capped at 1200 to 1320 words.

AI Interaction Logs: Every input/output exchange with the four constrained chatbots (outline, editing assistant, audience perspective, and devil's advocate) was automatically recorded. These logs generated a high-volume dataset, capturing hundreds of iterative prompts, tool responses, and follow-up queries per student. The result was a detailed record of student-AI engagement dense enough to support individual case tracking and broader analysis.

Student Reflections: Weekly reflections were collected via Mirror-Talk.ai, structured around the TRACE framework (Target, Refine, Assess, Cycle, Extend), and automatically transcribed by the system. Students reflected on tool effectiveness, rejected suggestions, ethical boundaries, and metacognitive awareness. Weekly reflections were 4–6 questions that required a 20 to 30 s recorded response each. Reflection

prompts focused on the targeted element for the week, such as: "During this week's exercises, how did working with AI examples change the way you think about adapting your message to different audiences? Can you walk me through a specific example?" Prompts addressed tool interaction: "What was most helpful about using AI tools for strengthening your arguments? What limitations did you encounter when working with AI?" Self-assessment: "Tell me one thing that you did well this week, and one thing that you still need to work on."

Semi-structured Interviews: Pre- and post-unit interviews lasting 45–60 min were conducted via Zoom, audio-recorded, and transcribed using Otter.ai. Interview protocols explored writing processes, AI attitudes, and perceptions of constraint-based design. Follow-up questions in the post interview clarified patterns observed in logs and writing samples. Transcripts were verified for accuracy before data analysis.

3.5. Data analysis

Metacognition was defined narrowly as observable self-regulatory behavior: monitoring, evaluating, or adjusting their AI tool use in relation to writing goals or ownership decisions [43]. Regulatory actions included rejecting AI feedback with articulated justification, describing the usefulness or limitations of a tool, modifying prompting strategies after unproductive feedback, and explaining decisions about when AI assistance supported or interfered with their work. Routine drafting, grammar or mechanics revision, and uncritical acceptance of AI output were treated as cognitive activity only. Appendix B provides a TRACE decision map linking each TRACE label to observable indicators, representative student evidence, and the argumentative dimensions summarized in Table 1.

Each TRACE label was applied using specific inclusion and exclusion criteria. *Target* was coded only when students named a precise writing goal ("I need to strengthen my counterclaim"), not for generalized improvement statements. *Refine* required students to revise or narrow their approach in response to feedback; surface-level edits or accepting AI suggestions without explanation were not coded. *Assess* was marked when students evaluated feedback or their own work ("I disagreed with

the AI because it contradicted my evidence”), not just when they accepted or ignored suggestions. *Cycle* required evidence of sustained, recursive revision attempts. *Extend* reflected students transferring strategies to new tasks or contexts. Borderline cases were resolved by returning to the full set of data and reviewing both context and prior coding decisions. Periodic review of logs, drafts, and reflections ensured TRACE-aligned routines remained present throughout the study, even though TRACE was not formally taught as a sequence.

Data analysis followed Braun and Clarke's [44] six-phase thematic analysis. A priori codes from the theoretical framework (counterargument development, self-regulatory behavior, ethical considerations) were applied across all data sources, then expanded through emergent coding. Initial analysis generated 25 codes that were consolidated through constant comparative analysis into three themes: (1) leveraging restricted AI feedback to refine writing processes, (2) developing strategic evaluation while upholding personal voice, and (3) constructing ethical boundaries through decisions about autonomy versus tool dependence. Triangulation across writing samples, interaction logs, reflections, and interviews revealed convergent patterns while preserving contradictions as evidence of complex adaptation processes.

Cognitive activity and metacognitive regulation were treated as separate forms of activity. Composing, revising text, and responding to feedback at face value were cognitive. Metacognitive regulation was coded only when students explicitly evaluated their own or the AI's contribution, justified acceptance or rejection of feedback, or adjusted feedback-seeking strategies based on assessed tool limitations or writing goals. Prompt refinement, for example, was cognitive when it reflected trial-and-error phrasing, and as metacognitive when students linked changes in prompting to evaluative judgments. To examine within-student change, initial and final drafts were compared with intermediate drafts to confirm sustained patterns of argumentation. Analysis was guided by established AP Seminar rubric constructs such as evidence integration, logical coherence, and counterargument development, which served as the analytic frame for classification. This approach reflects standard AP Capstone scoring practice, in which rubric fidelity anchors judgment. The first author's experience as an AP Seminar Table Leader informed consistent application of these constructs across drafts. Table 1 is not a rubric and does not represent mastery levels. Instead, classifications describe within-student change relative to each student's baseline.

Strengthened indicated clear improvement sustained across at least two intermediate drafts, with evidence that revisions addressed argumentative weaknesses (e.g., vague claims became focused, counterarguments moved from dismissive to substantive). *Partially Strengthened* indicated improvement that was present in some drafts but inconsistent across sections, limited in analytical depth despite structural gains, or evident in only one intermediate draft. *Not Strengthened* indicated no meaningful change or regression. Boundary decisions were resolved by returning to the full set of drafts, logs, and reflections to determine whether changes reflected strategic, evaluative decisions rather than simple expansion or compliance.

Given the intensive case-based design ($n = 4$), anchor examples in Appendix A represent the range of change observed rather than exhaustive coverage of all possible intermediate states. AI interaction logs and reflections were consulted repeatedly at multiple points in the analysis to verify that changes reflected evaluative decisions rather than routine revision.

3.6. Trustworthiness

Multiple strategies established trustworthiness [45]. Credibility was achieved through prolonged nine-week engagement, triangulation across four data sources, and systematic documentation of student writing, reflections, and AI interactions. Thick description of individual student trajectories supports transferability. Dependability was strengthened through detailed audit trails of the coding process and

constant comparative analysis. The first author's dual role as teacher-researcher created both insider knowledge of the instructional context and potential bias; co-authors with expertise in learning sciences and metacognition provided external perspective during analysis and interpretation phases.

4. Results

Participants demonstrated changes in argumentative reasoning and self-regulatory practices over the nine weeks of interaction with the constrained AI system. Three changes emerged: (1) counterargument development went from formulaic (or ignored) to genuine engagement with opposing views, (2) revision became iterative and substantive, and (3) students demonstrated increasing ability to craft targeted requests and evaluate AI feedback—prompt literacy—consistent with self-regulatory monitoring. Constraining AI outputs preserved thinking and provided strategic support.

4.1. Initial development and understanding

Students began with weak argumentative skills and shallow AI understanding. Their baseline essays lacked clear thesis statements, counterarguments, and source integration. Their AI experience was limited, seeing it as a shortcut for efficiency or an untrustworthy "cheating" tool. Luigi articulated the lack of focus on extended writing in school: "I feel like my essays improved up to a certain point, then I just kind of capped out. I think that was around seventh or eighth grade where I learned about counterarguments, and that's been the most recent thing I've learned that's new." His baseline reflected this, with loosely connected anecdotes, no clear thesis, no counterpoint, and no sources used for support.

By contrast, Katarina submitted the most coherent baseline, but her method remained source-by-source. Her comments reveal why: "[In middle school,] I used to get feedback like, 'this paragraph is off topic,' but no one ever really taught me how to fix it. I just sort of figured it out on my own." Her essay used effective strong citation practices, but the reasoning lacked integration; instead of contributing to a central claim, each paragraph stood on its own.

Similarly, other students showed source-by-source organization without synthesis or meandering arguments without structure. Over-reliance on summary, underuse of evidence, and absence of counterarguments was prevalent. In the pre-intervention interviews they described doing only minimal revisions. To construct effective arguments, students had to reconceptualize writing and AI as tools for inquiry, supported through targeted instruction and modeling. As the unit proceeded, the instructional boundaries supported student growth by requiring them to identify, evaluate, and revise their own arguments because the AI would not.

4.2. Constraints drive argumentative growth

The chatbot tools used early in the unit, particularly the outline and editing bots, encouraged clarification, specificity, and better logic. This restriction, while initially frustrating, drove students' argumentative development.

Betty spoke to this realization during her post interview. She admitted, "At first, I wouldn't really give it any direct task: I would just throw in my paper, and it wouldn't really give me any direct feedback." By the middle of the unit, she changed her prompting approach: "Then as we went on... I would like start with 'give me feedback on my conclusion or on my second body paragraph.'" This correction marked a turning point in her self-regulated learning development.

Katarina initially struggled with structural clarity. Her first outline resembled a narrative not a hierarchical academic outline that organized claims, evidence, and counterarguments. The chatbot did not offer fixes but pushed her to consider evidence placement and integrating ideas:

"How could I merge claim 1 and 2 about physical and mental health?" she asked, treating the bot as a brainstorming partner.

To support analytic transparency, Table 1 summarizes within-student change from baseline to final draft across six argumentative dimensions. Classifications (Strengthened/Partially Strengthened/Not Strengthened) are descriptive, relative to each student's baseline, and were confirmed by repeated comparison with intermediate drafts; AI logs and reflections were consulted to verify that revisions reflected evaluative decisions rather than simple expansion. The table shows that most participants demonstrated improvement or partial improvement in their ability to refine their overly broad claims, use evidence to support those claims, elaborate on why and how that evidence was relevant and organize their arguments into logical progression. Limits encouraged genuine revision, iterative improvements guided by students' evaluations. The chatbots helped students identify problems, but students decided which deserved attention and how to address them. In doing so, it also fostered their self-regulation as writers, requiring them to identify and articulate areas for improvement instead of just accepting "fixes".

4.3. Counterargument development and rebuttal growth

Students improved not only in their ability to introduce counterarguments but how they engaged with and rebutted opposing views, though still with room for development. The final two rows of Table 1 are the most interesting, highlighting how students moved from limited or absent acknowledgement of opposing perspectives to more deliberate, textually grounded engagement. The variation across students also showed that growth was not uniform. Some made more consistent gains across categories while others did not. These patterns suggest that constraint-based AI support can support more complex argumentative thinking and rhetoric, though growth depended on how students

engaged with the tool and the challenges and queries posed by the system Table 1.

"While you focus on the negative impacts of dependence on coaches, are there instances where this shift might lead to stronger relationships with mentors who provide a different type of support?" (counterpoint)

"What evidence might support the idea that an athlete can gain strengths from their experiences, even negative ones?" (Socratic)

"Explore literature on identity beyond sports; what psychological insights can be gained about multi-faceted identities in athletes?" (further investigation)

The structural changes were dramatic. Initially dismissive counterarguments: "This is incorrect because..." became thoughtful acknowledgements of opposition before deft rebuttals: "These are good points if you're thinking of the short-term, however, they become overshadowed by the physical remnants and behaviors that stick with an athlete, preventing full mental recovery." Instead of abrupt rejection, he brought in Jeong and Johnson to explain "long-term psychological damage, behavioral changes, and loss of identity post-injury," creating a layered argument that "walks through lasting effects." This supported the justification for the *Strengthened* classification in Table 1.

Luigi described this experience: "I'm using it less to give me ideas on what to write and more using it as a way to check if my work is sufficient." As the AI forced confrontation with uncertainty, students had to both add and reckon with counterarguments simultaneously.

4.4. From survival to strategic revision

Students learned that compliance with writing conventions did not

Table 1
Overview of argumentative growth.

Focus Area	Betty	Katarina	Michael	Luigi
Claim Refinement	Partially Strengthened - Expanded from vague to more specific ideas on FOMO.	Strengthened - Became more precise about smoking cessation's temporary vs. lasting effects.	Not Strengthened - Core claim about media's moral influence stayed broad.	Strengthened - Added emphasis on societal pressures & long-term injury impacts.
Evidence Use	Partially Strengthened - Integrated Winick & Cook but needed deeper analysis.	Strengthened - Used varied data on pandemic's effects with moderate synthesis.	Partially Strengthened - Cited Veera Ehrlén but shallow integration.	Strengthened - Reinforced arguments with Jeong & Johnson for mental health angles.
Reasoning Expansion	Partially Strengthened - Elaborated on FOMO's emotional vs. motivational aspects.	Partially Strengthened - Saw bias in personal sources, tried to broaden logic but still brief.	Partially Strengthened - Discussed media-induced confusion but lacked direct link to moral reasoning.	Strengthened - Contrasted short-term gains with long-term injuries.
Logical Flow	Partially Strengthened - Transitions improved but still choppy in places.	Strengthened - Linked short-term pandemic changes to broader health patterns.	Partially Strengthened - Minor transitions improved but abrupt shifts remain.	Partially Strengthened - Distinguished short- vs. long-term effects, though some abruptness.
Counterargument & Rebuttal	Strengthened - Added positivity vs. anxiety contradiction for FOMO.	Strengthened - Countered idea that quitting = lasting health improvements.	Partially Strengthened - Rebuttal present but under-supported.	Strengthened - Emphasized long-term negative impacts to rebut 'injuries foster only growth.'
Perspectives in Conversation	Strengthened - Referenced multiple authors (Winick & Cook) in dialogue.	Not Strengthened - Hoo & Pierce used but not truly woven together.	Not Strengthened - Cited Ehrlén but did not expand other viewpoints.	Strengthened - Linked Trainor & Johnson's viewpoints on short- vs. long-term injury effects.

Note: Cells with a white background demonstrated improvement, light grey showed partial improvement, and dark grey showed no improvement.

Note: Table shading indicates degree of change (Strengthened/Partially Strengthened/Not Strengthened). Luigi's growth in counterargument development shows how evidence supporting Table 1 classifications were identified and interpreted. In early drafts, he asserted the psychological harms of sports injuries without acknowledging alternate perspectives. The devil's advocate prompted him to engage with structured counterpoint, Socratic, and further investigation questioning, leading him to consider athlete resilience research.

equal engagement and adjusted accordingly. Initially, Michael described his approach to AI as a workaround for high stakes, time intensive writing tasks: "I'm buying my freedom for time so I can do what I want later," in an academic structure that deprioritized revision: "There's no revision, just hopefully do better." This efficiency-seeking behavior reflected performance goals, completing tasks to avoid failure, rather than engaging with learning processes.

The constrained AI environment disrupted this pattern. Early frustration led Michael to develop boundaries: asking the chatbot for three-sentence summaries when extensive feedback felt "too open." He was monitoring how he processed feedback (summaries) and deliberately adjusting the format so it matched the way he could think through it to calibrate feedback he could actually use.

His post interview reflection noted: "If you don't understand the answer, you might feel stuck. But I asked it [the bot] to explain further, and it gave a better explanation. So, pushing back when I didn't understand was very vital." By the end, AI became his sparring partner: "It's only supposed to be a helper and not become your voice."

Students came to the realization that revision was something *worth doing*. Luigi, who initially relied on the organizational structure suggested by the bot, learned when to step back. "I learned that even though AI is a tool, I have to do the thinking to make sure it's actually helping me." Betty's evolution followed suit. Her early reflections: "I just threw in my paper and hoped it told me what was wrong." But devil's advocate interactions helped her see revision as "an argument with someone else": "I used to think arguments were hard but now I know when you see other people's perspectives, it's easier to create an effective argument." She was actively anticipating how different audiences would respond and using those reactions to strengthen and test her own claims.

Across all students, revision was no longer reactive but anticipated, because strategic AI use let them reject unhelpful suggestions that obscured their style. Students showed observable self-regulatory behavior, self-monitoring, and intentional planning, supported by strategic tool use and growing prompt literacy. Early prompts were broad or vague, producing unfocused feedback. Students refined their requests, moving from "check my paper" to "How is this specific evidence in line with my claim?"

4.5. Learning to ask better questions

Prompting through better questioning was a topic of frequent discussion and an immutable component of the scaffold. By Week Four, learning to craft targeted questions became a metacognitive skill guided by clearer thinking about their own writing needs, extending beyond AI use.

Prompt evolution directly reflected increasingly strategic regulation of feedback-seeking behavior. Michael initially struggled: "When you give the AI, like, too broad of a question, like, you have to phrase your question in a way that will give you the answer that you want." When he felt "buried by feedback," he learned to reject suggestions mid-thread: "No thank you. How do I construct a good 1–2 sentence conclusion for my solutions and limitations?" By the unit's end: "Over time, it became stronger, because I knew what answer I was looking for."

Betty learned to craft task-oriented prompts: "I would start with like, 'Give me feedback on my conclusion or my second body paragraph.' I gave better instructions."

Katarina's understanding of how tool limitations affect outputs was particularly insightful. Her early prompts led to unusable suggestions: "With some of the suggestions, it would say like... tie in another lens here where it didn't necessarily make sense to." Her eventual solution came from improved grasp of rhetorical context with her realization that the tools didn't understand her rhetorical context: "This editing tool only has the information that we have given it. It doesn't have access to everything online, so I can't be broad. I have to be specific."

Students increasingly attributed unhelpful outputs to either imprecise prompts or tool limits and adjusted their requests accordingly. In

turn, this made students better self-diagnostics. Identifying needed revisions became a habit in self-regulated cycles: simultaneously composing and evaluating, considering how to improve, clarify, and personalize their work into their writing process by unit's end. This constitutes metacognitive regulation; students recognized that the AI tools had limits, evaluated any mismatch between output and intent, and adjusted their strategies deliberately to better align feedback with writing goals.

4.6. Ethical boundaries through design

The constrained environment supported moral reasoning that emerged from repeated decisions about tool dependence and authorship when writing. In this study, "evaluative ethics" refers to students' observable decision-making about when AI assistance supported their goals and when it threatened authorship or voice, as expressed through tool rejection, modification, or justification. Unlike rule-based approaches to AI ethics, students developed personal boundaries and made ethical judgements through lived experience with tools that demanded ongoing ethical negotiation.

Luigi's evolution illustrates this process. Initially tempted to copy AI suggestions wholesale, he learned to distinguish between clarification and replacement: "Eventually I would ask it, oh, can you better clarify the sentence, but I still had to, like, redo that sentence." This was not compliance but recognition that authorship required active choice-making each time.

Katarina developed the most explicit ethical boundary, drawing clear distinctions between acceptable and unacceptable assistance: "I decided not to use features like sentence-flipping because I felt it crossed ethical boundaries... It wasn't my words." She consistently "kept her language, rejecting anything too polished or robotic," showing that ethical reasoning extended beyond plagiarism concerns to questions of authentic voice.

These patterns suggest that ethical judgment becomes more visible when tool design requires repeated decisions about ownership and dependence.

5. Discussion

This design-based study examined how a coherent instructional system functioned in practice. Constrained tools, TRACE framework, and teaching routines were intentionally interdependent, reflecting how technology is implemented in working classrooms. Designing strategic constraints encouraged cognitive engagement instead of cognitive off-loading [26], challenging assumptions that AI tools promote cognitive laziness. Three theoretically distinct but interrelated phenomena emerged. *Constraint-as-affordance*, the design principle, preserved argumentative reasoning by limiting AI output to questioning to make certain cognitive processes unavoidable [23]. This elicited *Productive friction*; students' phenomenological experience of persistent resistance when tools refuse to complete their thinking. Friction remained constant throughout the unit even as students gained access to more tools, because prompt limits tightened weekly, requiring increasingly strategic and economical tool use. Students did not find the tools less annoying over time, they developed competence in converting sustained resistance into argumentative work by developing "prompt literacy" as a new metacognitive competence for strategic human-AI collaboration. *Evaluative ethics* emerged as students repeatedly navigated this friction, developing judgment about when AI assistance supported their goals versus when it threatened authorship. The theoretical sequence matters: architectural constraints produce experiential friction, which drives regulatory development. The constraint stayed firm while student capability grew, which enabled the self-regulatory development we observed.

5.1. Scaffolded constraints and argumentative development

Gibson's ecological approach to affordance theory frames how limited-output AI design enhances cognitive development through relationships between tools, environments, and learner capabilities [23, 24]. The AI tools afforded specific cognitive actions while making others difficult or impossible and prompted critical examination by disrupting certainty, but only if students engaged with that resistance.

The AI design in this study paralleled effective instructional practices for writing development. Effective writing instruction engages students through strategic questioning that guides thinking, much like the constrained AI tools refused to complete students' thinking. However, implementing these strategies consistently in traditional classrooms proves challenging, particularly when students resist cognitive effort or teachers with large classes are short on time. The constraint in the scaffold was inconvenient but maintained the consistent approach that effective human instructors can struggle to sustain. Students had to draft before tool use, could not use full generative outputs, and were coached to interpret and challenge AI prompts. This preserved the effortful processing that drives argumentative development, aligning with cognitive apprenticeship principles of gradual release [40,46]. Students had to stay active even with support, which aligns with research on desirable difficulties, showing strategic obstacles can enhance learning [25,19].

Repositioning from performance-oriented to mastery-focused engagement was significant for revision. Students viewed revision as necessary for compelling arguments. This change aligns with achievement goal theory, where mastery goals foster more engagement than performance goals [47]. The constrained AI environment disrupted efficiency-seeking behaviors by making cognitive shortcuts impossible, redirecting student attention from task completion to process improvement. This mirrors novice-to-expert argumentative development [13, 34]. This resistance set the stage for self-regulation, as students worked through the difficulty.

5.2. Self-regulation development through prompt literacy

The TRACE model (Target, Refine, Assess, Cycle, Extend) supported self-regulation, but student development occurred through repeated AI feedback requiring judgment in embedded practice. Detached instruction does not lead to self-regulation; embedded practice does [48–50]. Students learned to identify when they needed help (Target), craft specific requests (Refine), evaluate usefulness (Assess), adjust their approach (Cycle), and apply insights to their writing (Extend).

The transition students described—from experiencing the AI as frustrating or "annoying" to viewing it as a helpful "sparring partner"—appears to have been mediated by several interacting factors that we cannot fully disambiguate. First, *increased task clarity*: students learned to diagnose specific weaknesses in their writing, which enabled more targeted prompting (Betty: "I would start with 'give me feedback on my conclusion or my second body paragraph'"). Second, *reduced automation bias*: repeated non-generative responses disrupted expectations that AI would provide direct answers, forcing students to evaluate rather than accept (Michael: "pushing back when I didn't understand was very vital"). Third, *pedagogical normalization of struggle*: classroom instruction, modeling, and reflection prompts consistently framed resistance as productive rather than obstructive, making friction interpretable as a feature rather than a bug. Fourth, *growing evaluative confidence*: early successes with rejecting, refining, or redirecting AI feedback built students' belief that they could manage the tool strategically rather than being managed by it.

These factors likely worked in concert rather than sequentially. The tool design created persistent resistance, but the instructional context—teacher modeling, reflection requirements, explicit discussion of authorship boundaries—made that resistance productive. This suggests that constraint-based AI functions best when embedded in a pedagogical

contract where struggle is expected, supported, and reframed as learning rather than failure. Without that framing, question-only AI might simply frustrate students into disengagement or workarounds.

Evaluative resistance was an important component of the TRACE model; students learned to say no. When AI suggestions failed, they pushed back. By the midpoint, students modified prompts, challenged AI suggestions, and rejected outputs that did not align with their goals. The change students discussed from experiencing the AI as frustrating or "annoying" to helpful appears to have been mediated by three interacting factors: increased task clarity, as students learned to diagnose specific weaknesses in their writing; reduced automation bias, as repeated non-generative responses disrupted expectations of direct answers; and growing evaluative confidence, as students experienced success rejecting, refining, or redirecting AI feedback. Unlike Darvishi et al. [31], which suggest reliance on AI without learning feedback strategies, our results show evidence of consistent, internalized output evaluation. These skills and habits may constitute a new literacy domain combining traditional critical thinking with an understanding of AI capabilities and limitations.

In contrast to recent multimodal studies [27,28], which frame prompting as expansive and playful, supporting invention, multimodal composing, and creative proliferation, our findings show that argumentation benefited from constraint. Initially, our students experimented with chatbot limits and testing restrictions, but as the unit progressed, their focus shifted toward narrowing claims, rejecting unhelpful feedback, and sustaining argumentative focus. This suggests the value of AI is genre dependent. This aligns with Smith et al.'s [28] argument that prompting can be seen as a literacy, but our evidence reframes it as regulatory practice, not iterative play. Over time, early trial-and-error manipulation of the tools developed into more expert-like behavior, paralleling self-regulatory cycles in expert learning [43,51]. As AI becomes commonplace in education, prompt literacy could become as fundamental as digital literacy was for previous generations [52], especially guiding decisions about the ethical use of AI. This raises the harder question: not just how students think with AI, but how they decide if they should.

5.3. Ethical reasoning through environmental design

The constrained AI environment elicited evaluative ethical reasoning, evidenced through students' repeated, situated decisions about authorship, dependence, and voice during tool use. Our findings extend Ranade and Eyman's [26] call for critical engagement with generative AI. Students developed evaluative ethical boundaries through repeated lived dilemmas, a form of "restriction-driven ethics" that contrasts with frameworks positioning AI as an enabler of creative choice. It aligns with Rath's [53] values clarification theory, emphasizing ethical development through actively making choices, and Rest's [54] model of moral judgment formation through real-world experience. Our students demonstrated this pattern, developing sophisticated judgments about when and how to use AI assistance. They learned their AI knowledge was incomplete, and tools are fallible and need guardrails. Boundaries emerged from experience, not policy, important distinction when over-reliance on AI can erode self-sufficiency and confidence in academic work (Rodríguez [55]) and risk leeching creativity and analytical ability from writers [56]. Unlike rule-based AI ethics, students developed personal boundaries and demanded ongoing ethical negotiation [57], because the tools refused to produce complete text. Deference to feedback lessened through engagement with teacher and bot feedback, disrupting the automation bias described by Mosier and Skitka [20]. When tools refused to complete students' thinking, every interaction became an opportunity to negotiate where assistance ended and intellectual encroachment began.

Users develop ethical positions through ongoing tool interaction instead of just compliance with policy [58]. AI users must navigate opaque algorithms, privacy risks, and human-AI authority without clear

guidance, creating opportunities for ethical sense-making [59]. Static rules gave way to situational evaluation: students evaluated AI suggestions situationally by testing, rejecting, or adapting outputs that reflected growing ethical judgment.

These findings affirm and complicate the current discourse. Prompting is now central to composing [27,28], and AI use in writing instruction is intertwined with ethical reflection, even if classrooms treat them as separate. But unlike studies highlighting AI's potential, our evidence demonstrates the value of narrowing options to make AI "annoying", so students sustain the intellectual work of argumentation.

5.4. Limitations

This study was conducted with a small group of highly capable AP Capstone students in a single instructional context. The instructional design combined constrained AI tools with sustained teacher guidance and structured reflection, so observed changes should be interpreted as outcomes of the system rather than attributed to any single design element. The data intensity was necessary to establish the viability of constraint-based design and document its effects on argumentative development; instructional implementation would not require this same depth of analysis.

The study did not examine transfer beyond constrained AI-supported writing tasks. Whether students apply similar evaluative strategies when using unrestricted AI tools or in non-AI writing contexts is unclear. Ethical reasoning was examined only within constrained AI-supported writing tasks.

5.5. Implications for writing instruction

The approach tested in this study suggests changes in how writing instructors approach AI integration. AI capabilities are not inherently beneficial or problematic; these findings point toward pedagogical strategies that preserve the cognitive work of argumentation by structuring how, when, and why students interact with AI [60,61].

The workload challenges of writing instruction, including reading multiple drafts, providing actionable feedback, and tracking student progress, are well-documented constraints that limit how much process-oriented teaching occurs in practice. In this study's context, AP Capstone policies prohibiting individual feedback after the first semester created conditions where constrained AI filled a necessary instructional gap. However, this gap is not unique to AP Capstone. In undergraduate composition courses and many high school settings, instructors managing large course loads struggle to provide the kind of iterative, individualized feedback that supports argumentative development [5].

Strategic use of constrained AI tools may offer a partial solution by redistributing feedback labor: bots can handle routine feedback interactions while instructors focus on higher-level process assessment. When AI tools provide consistent questioning and students document their interactions through reflections, instructors gain consolidated artifacts that could surface learning patterns more efficiently than manual draft-by-draft comparison. However, this potential requires empirical validation and likely depends on learning analytics or AI-assisted summarization to make pattern detection tractable at scale. The present study cannot speak to workload implications beyond demonstrating that constraint-based design functions in one well-supported classroom context.

5.5.1. Classroom strategies

Teachers can encourage strategic AI interaction through low-stakes diagnostic tasks and flexible grouping. Students who learned to craft targeted questions based on self-assessment developed stronger self-regulatory awareness than those relying on broad AI assistance [62, 63]. Scaffolding can be differentiated: some might work with structured prompt templates, while advanced writers critique and refine AI output. Targeted mini lessons can model how to identify a writing weakness,

craft a precise prompt, and evaluate an AI response. The goal is not initial efficiency, but for students to develop prompts aligned with their argumentative sophistication (Stojković & Wang, 2024). Recognizing when students are ready for reduced restrictions is necessary. Students ready for fewer constraints typically demonstrate three behaviors: precision in feedback requests, consistent evaluation of AI suggestions, and articulated ethical boundaries for AI use (Tankelevitch et al., 2023). As students demonstrate these behaviors, they can access more generative AI features while retaining ownership over their writing process [60]. Quick one-on-one or small group "prompt conferences" can help teachers assess readiness.

5.5.2. Assessment and ethical development

Assessment practices could adapt to document process alongside the final product. Evidence of the process can be captured through portfolios documenting AI interactions, rejected suggestions, and reflections on tool use [64]. Whether such documentation is feasible without analytic support in large-enrollment contexts remains an open question. Furthermore, ethical reasoning should not be treated as a detached set of classroom rules, but as something cultivated through the instructional design [58,65]. When tools resist overreach, students are pushed into their own authorship, making choices that embed ethical judgment into the writing process.

5.5.3. Institutional support

Supporting these classroom changes requires institutional commitment. Teachers and students need time to experiment before formal scaffolding begins. Early, low stakes "play" with AI systems helps participants build mental models of these tools. Therefore, institutional support should prioritize creating protected time and space for exploration, in addition to providing formal professional development opportunities.

6. Conclusion

Every wave of technology tries to lighten the mechanical burdens of a craft, freeing people to focus more intensely on developing ideas; writing is no exception. Word processors spared us the mess of ink, citation managers kept our footnotes in line, but the *work of thought* remained the writers' domain. Generative AI can cross that line, able to finish a thought before a writer has a chance to fully form it. A student can submit a neat and seemingly substantive essay without ever having to face the slow, uncomfortable labor that turns words into understanding.

Ethical reasoning about AI use need not rely on abstract policies or honor codes. It can emerge from the material conditions of tool interaction itself. Students developed "evaluative ethics": situated judgment about when assistance supports versus threatens authorship, formed through repeated decisions about dependence and autonomy during everyday writing work. When AI tools refuse to complete students' thinking, every exchange becomes an occasion for negotiating intellectual ownership. This stands in contrast to rule-based approaches where students may comply under supervision but default to convenience under pressure. By making ethical decisions unavoidable through design, constraint-based AI embeds moral reasoning into the writing process rather than treating it as a separate instructional concern.

Writing instruction has always focused more on process than product. Accordingly, the instructional design principles from our study offer a framework for teaching argumentative writing. Making AI "annoying" on purpose kept students in the hard parts of thinking with tools that made students better at the work.

CRedit authorship contribution statement

Lindsay Konradt: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data

curation, Conceptualization. **David N. Boote:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis. **Michelle Taub:** Writing – review & editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgements

I would like to thank the members of my dissertation committee, Dr. Ceballos, and Dr. Campbell, for their guidance and support. I am also deeply grateful to the students who participated; their time, insights, and willingness to share made this research possible. Article processing charges were provided in part by the UCF College of Graduate Studies Open Access Publishing Fund.

Appendix A. Representative initial and revised examples Table 1

Analytic Category	Student	Initial draft	Revised draft
<i>Claim Refinement</i>	Luigi	"Sports injuries really affect athletes mentally and physically and can change their lives forever."	"These are really good points if you're thinking of the short-term, however, they become overshadowed by the physical remnants and behaviors that stick with an athlete, preventing full mental recovery."
<i>Evidence Use and Integration</i>	Betty	"Winick says that FOMO does not affect teenagers."	"Winick and Cook note that FOMO can inspire action but also cause anxiety."
<i>Reasoning Expansion</i>	Betty	"FOMO affects teenagers because they feel left out."	"When teenagers view social media, they may feel like they don't have such a good life or that they are doing something wrong, which can increase anxiety."
<i>Counterargument and Rebuttal</i>	Luigi	"This, however, is incorrect because injuries negatively affect mental health."	"Psychologist Lauren Jeong explains that if an athlete gains negative behavioral traits during adolescence, they can carry on into adulthood... The buildup of these changes in one's behavior completely alters who the person is."
<i>Logical Flow</i>	Michael	"Social media affects people in many ways. It can be helpful but also harmful. People use it for communication, but it also causes problems."	"Although social media allows for increased communication, its influence on moral development becomes problematic when repeated exposure to conflicting values leads to confusion about ethical norms."
<i>Perspectives in Conversation</i>	Betty	"Winick explains how FOMO affects teenagers."	"These authors both support that FOMO from social media can have a positive influence but Winick explains how the fear of missing out can spark anxiety and unhappiness."

Appendix B. TRACE decision map aligned with Table 1

TRACE Label	Observable Indicator	Representative Student Quote	Data Source	Table 1 Classification
<i>Target</i>	Student identifies a specific argumentative problem prior to revising or seeking feedback	"I need to make sure that my counterargument is strong to rebut my point and really understand the other point of view." (Betty)	Weekly reflection	Counterargument & Rebuttal; Claim Refinement
<i>Target</i>	Student identifies a specific argumentative problem prior to revising or seeking feedback	"Short doesn't really mean better. I realized it just means I don't have enough support." (Luigi)	Weekly reflection	Evidence Use; Reasoning Expansion
<i>Refine</i>	Student revises or narrows feedback requests after recognizing that broad prompts are ineffective	"If I ask it everything at once, it gives me way too much. I started only asking about one thing at a time." (Katarina)	Weekly reflection	Logical Flow; Reasoning Expansion
<i>Refine</i>	Student revises or narrows feedback requests after recognizing that broad prompts are ineffective	"When I made the questions more specific, it stopped just repeating stuff I already knew." (Luigi)	Weekly reflection	Claim Refinement; Evidence Use
<i>Assess</i>	Student evaluates AI feedback against task constraints, goals, or genre expectations	"A lot of the stuff it was telling me, like using first person in my paper, is physically something I cannot use." (Katarina)	Post-unit interview	Logical Flow
<i>Assess</i>	Student evaluates AI feedback against task constraints, goals, or genre expectations	"I didn't just use what it said. I had to decide if it actually fit what we were allowed to do." (Betty)	Weekly reflection	Reasoning Expansion
<i>Cycle</i>	Student adjusts prompting or revision strategy based on evaluation of prior feedback	"At first I tried fixing everything it suggested, but that didn't work. I started picking what actually mattered." (Michael)	Weekly reflection	Logical Flow; Evidence Use
<i>Cycle</i>	Student adjusts prompting or revision strategy based on evaluation of prior feedback	"I used to just avoid revising because it felt pointless, but breaking it down made it easier to actually fix things." (Luigi)	Post-unit interview	Claim Refinement; Logical Flow
<i>Extend</i>	Student independently applies previously evaluated strategies in later drafting or planning	"Using AI tools has changed the way I construct my essays. Now I try to think about who I'm writing for before I even start." (Michael)	Weekly reflection (Week 8)	Perspectives in Conversation; Reasoning Expansion; Counterargument & Rebuttal
<i>Extend</i>	Student independently applies previously evaluated strategies in later drafting or planning	"I don't just write everything out anymore. I organize my claims first so the evidence actually matches." (Luigi)	Final reflection	Claim Refinement; Evidence Use

Appendix C. System Prompts, Refusal Behavior, Escalation Protocol, and Constraint Parameters

System Prompts for Constrained AI Tools

All chatbots in this study used a shared system-level instruction, with additional tool-specific guidance layered on top. The base system prompt was:

“Your role is to support argumentative writing through questioning only. For every user input, respond with 2–3 targeted, content-specific questions designed to prompt further reasoning or clarification. Do not provide direct answers, solutions, claims, text corrections, or examples of improved writing. If a user requests you to write, rewrite, generate text, or answer directly, politely refuse and reiterate your role as a questioning partner.”

Tool-Specific Rule examples

Devil’s Advocate Bot: Your questions should focus on challenging claims, surfacing counterarguments, and probing evidence or assumptions.

Audience Perspective Bot: Your questions should address rhetorical context, audience values, and likely reactions.

Outline Bot: Your questions should probe logical order, argument structure, and relationships between claims.

Editing Assistant: Your questions should focus on clarity, coherence, and genre expectations—never rewrite or suggest alternative sentences.

Examples of Refusal Behavior

If prompted with: “Can you rewrite my introduction?”

Response: “I’m here to help by asking questions, not by rewriting text. What goal do you have for improving your introduction?”

If prompted with: “What’s a good way to phrase this argument?”

Response: “I can help you think through your argument by asking questions. What claim are you trying to make, and what evidence supports it?”

If asked for direct corrections or suggestions:

Response: “I cannot correct or edit your text directly. Instead, consider what you want your audience to understand, and let’s explore that through questions.”

Escalation Protocol: Adaptive Scaffolding When Students Stalled

If a student indicated confusion, lack of direction, or persistent non-responsiveness (e.g., “I don’t know what to do next” or repeated prompts with no substantive change), the escalation protocol was as follows:

1. First escalation: The AI would repeat its questioning function, offering more specific or guided questions such as, “What’s one area you’re most uncertain about in your argument?”
2. Second escalation: If the student continued to struggle, I would intervene directly—through modeling, targeted live demonstration, or small-group guided practice. At no point did the AI generate claims, prose, or solutions.
3. Documentation: All escalations were logged. If a student made three or more consecutive unsuccessful attempts, I reviewed the log to determine if additional instructional scaffolding was needed.

AI Chatbot Constraint Parameters

Parameter	Implementation
Input requirements	Open text from student (draft, outline, paragraph, or free-response prompt)
Refusal triggers	Any request to write, rewrite, generate, correct, or answer directly
Output format	2–3 numbered, content-specific questions per response
Max turns per session	Reduced weekly
Question categories	Counterpoint, Socratic, Further Investigation (mapped to Table 1 analytic categories)
Personalization	No adaptation to student proficiency or interaction history; system was not adaptive
Escalation path	Repeated questions to more targeted prompts to teacher intervention if needed
Log retention	All student-AI exchanges automatically recorded and available for analysis or instructional review

Appendix D. TRACE coding rules, coded examples, and fidelity verification

TRACE Label	Inclusion Rule	Borderline/Excluded Cases	Typical Data Source
Target	Student explicitly states a specific writing need, weakness, or goal before acting or seeking feedback (e.g., “I need to add a counterargument to this paragraph.”)	Vague statements (“I need to make this better”) without reference to a particular claim, evidence, or argument. Generalized frustration without a plan.	Reflection, log, draft margin comment
Refine	Student revises a prompt or narrows a request in response to prior feedback (“Last time I asked about everything, now I’m just focusing on my evidence.”)	Prompt change with no rationale or strategic focus. Accidental prompt rewording.	Reflection, AI log
Assess	Student evaluates feedback against task demands, genre norms, or personal goals (“That suggestion won’t work for AP Seminar because we can’t use first person.”)	Passive acceptance/rejection with no reason (“I didn’t like it”). Routine compliance without evaluation.	Reflection, interview
Cycle	Student adjusts writing or prompting strategy <i>based on prior assessment</i> (“I fixed the transition because last time the bot said it was abrupt.”)	Multiple drafts with changes but no linkage to prior evaluation; mechanical revision.	Drafts, reflection, log
Extend	Student applies a previously learned/evaluated strategy to a new section, context, or paper (“Now I organize my claims before starting, just like with my last essay.”)	Repetition of a strategy in the same context without awareness (“I just kept revising it again and again.”)	Reflection, final interview

Coded Examples: Inclusion and Exclusion

Example: Target

Included:

"I realized my rebuttal was weak, so I tried to focus my next draft on explaining the other side better." (Coded: Target; Specific weakness identified and intent to address)

Excluded:

"I wanted to improve my essay." (Not coded: Too vague; no specific goal or area targeted)

Example: Assess

Included:

"The bot told me to use first person, but that's not allowed for this assignment." (Coded: Assess; Evaluation against task constraints)

Excluded:

"I didn't use the suggestion." (Not coded: No reason given for rejecting feedback)

Fidelity Verification: How TRACE-Aligned Routines Were Ensured

I kept a checklist during each instructional session of modeled TRACE-aligned moves (e.g., explicit targeting of a weakness, modeling prompt refinement, pausing to assess AI feedback, etc.).

I reviewed student logs and reflections each week to verify that TRACE-aligned behaviors appeared in each cycle (targeting, refining, assessing, cycling, or extending).

Fidelity was monitored by aligning every lesson plan, demonstration, and reflection prompt to at least one TRACE category. Any deviation was recorded and addressed in subsequent sessions.

Student products and process artifacts (logs, drafts, reflections) were regularly mapped to the TRACE framework as a form of process audit, with representative exemplars included in [Appendix B](#).

During the preparation of this work the author(s) used Lex.page in order to assist with outlining and improving the manuscript's conciseness. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- [1] Liu F, Stapleton P. Counterargument and the cultivation of critical thinking in argumentative writing: investigating washback from a high-stakes test. *System* 2014;45:117–28. <https://doi.org/10.1016/j.system.2014.05.005>.
- [2] Wolfe CR, Britt MA, Butler JA. Argumentation schema and the myside bias in written argumentation. *Writ Commun* 2008;26(2):183–209. <https://doi.org/10.1177/0741088309333019>.
- [3] Kuhn D, Udell W. The development of argument skills. *Child Dev* 2003;74(5):1245–60. <https://doi.org/10.1111/1467-8624.00605>.
- [4] Applebee AN, Langer JA. A snapshot of writing instruction in middle schools and high schools. *Engl J* 2011;100(6):14–27. <https://doi.org/10.58680/ej201116413>.
- [5] Graham S. Changing how writing is taught. *Rev Res Educ* 2019;43(1):277–303. <https://doi.org/10.3102/0091732X18821125>.
- [6] Zimmerman BJ, Moylan AR. Self-regulation: where metacognition and motivation intersect. In: Hacker DJ, Dunlosky J, Graesser AC, editors. *Handbook of metacognition in education*. Routledge; 2009. p. 299–315.
- [7] Butler AC, Britt MA. Investigating instruction for improving revision of argumentative essays. *Writ Commun* 2011;28(1):70–96. <https://doi.org/10.1177/0741088310387891>.
- [8] Freedman SW, Daiute C. *Instructional methods and learning activities in teaching writing*. 8. Emerald Group Publishing Limited; 2001. [https://doi.org/10.1016/S1479-3687\(01\)80024-9](https://doi.org/10.1016/S1479-3687(01)80024-9).
- [9] Selfe CL. *Technology and literacy in the twenty-first century: the importance of paying attention*. Southern Illinois University Press; 1999.
- [10] Aljuaid H. The impact of artificial intelligence tools on academic writing instruction in higher education: a systematic review. *Arab World Engl J* 2024. <https://doi.org/10.31235/osf.io/ph24v>.
- [11] Holstein K, McLaren BM, Aleven V. Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In: *Proceedings of the 7th International Conference on Learning Analytics & Knowledge*; 2019. p. 257–66. <https://doi.org/10.1145/3027385.3027451>.
- [12] Mogavi, R.H., Deng, C., Kim, J.J., Zhou, P., Kwon, Y.D., Metwally, A.H.S., & Hui, P. (2023). Exploring user perspectives on chatgpt: applications, perceptions, and implications for AI-integrated education. *arXiv preprint arXiv:2305.13114*. <https://doi.org/10.48550/arXiv.2305.13114>.
- [13] Bereiter C, Scardamalia M. *The psychology of written composition*. Routledge; 2013.
- [14] Selber SA. *Multiliteracies for a digital age*. Southern Illinois University Press; 2004.
- [15] Popenici SA, Kerr S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res Pract Technol Enhanc Learn* 2017;12(1):22. <https://doi.org/10.1186/s41039-017-0062-8>.
- [16] Selwyn N. The future of AI and education: some cautionary notes. *Eur J Educ* 2022; 57(4):620–31. <https://doi.org/10.1111/ejed.12532>.
- [17] Göldi A, Wambsgans T, Neshaei SP, Rietsche R. Intelligent support engages writers through relevant cognitive processes. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*; 2024. p. 1–12. <https://doi.org/10.24451/arbort.22084>.
- [18] Risko EF, Gilbert SJ. Cognitive offloading. *Trends Cogn Sci (Regul Ed)* 2016;20(9):676–88.
- [19] Sweller J, van Merriënboer JJ, Paas FG. Cognitive architecture and instructional design. *Educ Psychol Rev* 1998;10(3):251–96. <https://doi.org/10.1023/A:1022193728205>.
- [20] Mosier KL, Skitka LJ. *Human decision makers and automated decision aids: made for each other?*. Automation and human performance. CRC Press; 2018. p. 201–20.
- [21] Horowitz MC, Kahn L. Bending the automation bias curve: a study of human and AI-based decision making in national security contexts. *Int Stud Q* 2024;68(2). <https://doi.org/10.1093/isq/sqae020>.
- [22] Fan Y, Tang L, Le H, Shen K, Tan S, Zhao TS, Shen Y, Li X, Gašević D. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance. *Brit J Educ Technol* 2024. <https://doi.org/10.1111/bjet.13544>.
- [23] Gibson JJ. *The ecological approach to visual perception*. Lawrence Erlbaum Associates; 1986.
- [24] Greeno JG. Gibson's affordances. *Psychol Rev* 1994;101(2):336–42. <https://doi.org/10.1037/0033-295X.101.2.336>.
- [25] Bjork RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe J, Shimamura AP, editors. *Metacognition: knowing about knowing*. MIT Press; 1994. p. 185–205.
- [26] Ranade N, Eyman D. Introduction: composing with generative AI. *Comput Compos* 2024;71. <https://doi.org/10.1016/j.compcom.2024.102834>.
- [27] Jiang J. When generative artificial intelligence meets multimodal composition: rethinking the composition process through an AI-assisted design project. *Comput Compos* 2024;72. <https://doi.org/10.1016/j.compcom.2024.102883>.
- [28] Smith BE, Shimizu AY, Burriss SK, Hundley M, Pendergrass E. Multimodal composing with generative AI: examining preservice teachers' processes and perspectives. *Comput Compos* 2024;75. <https://doi.org/10.1016/j.compcom.2024.102896>.
- [29] Delfavero, L., Pérez-Ortiz, J.A., Käser, T., & Oliver, N. (2024). Enhancing critical thinking in education by means of a socratic chatbot. <https://doi.org/10.48550/arxiv.2409.05511>.
- [30] Afrin T, Kashefi O, Olshefski C, Litman DJ, Hwa R, Godley A. Effective interfaces for student-driven revision sessions for argumentative writing. *Hum Fact Comput Syst* 2021. <https://doi.org/10.1145/3411764.3445683>.
- [31] Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. *Comput Educ* 2024;210. <https://doi.org/10.1016/j.compedu.2023.104967>.
- [32] Cuban L. *Oversold and underused: computers in the classroom*. Cambridge, MA and London, England: Harvard University Press; 2001. <https://doi.org/10.4159/9780674030107>.
- [33] Graham S, Perin D. A meta-analysis of writing instruction for adolescent students. *J Educ Psychol* 2007;99(3):445. <https://doi.org/10.1037/0022-0663.99.3.445>.
- [34] Flower L, Hayes JR. A cognitive process theory of writing. *Coll Compos Commun* 1981;32(4):365–87. <https://doi.org/10.2307/356600>.
- [35] Hillocks G. Research on written composition: new directions for teaching. In: *National Conference on Research in English*; 1986. <https://doi.org/10.2307/357721>.

- [36] Barab S, Squire K. Design-based research: putting a stake in the ground. *J Learn Sci* 2004;13(1):1–14. https://doi.org/10.1207/s15327809jls1301_1.
- [37] Harris KR, Graham S. Self-regulated strategy development: theoretical bases, critical instructional elements, and future research. *Design principles for teaching effective writing*. Brill; 2017. p. 119–51.
- [38] Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn Environ* 2024;11(1). <https://doi.org/10.1186/s40561-024-00316-7>.
- [39] Cukurova M, Kent C, Luckin R. Artificial intelligence and multimodal data in the service of human decision-making: a case study in debate tutoring. *Brit J Educ Technol* 2019;50(6):3032–46. <https://doi.org/10.1111/bjet.12829>.
- [40] Collins A, Brown JS, Newman SE. Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. *Knowing, learning, and instruction*. Routledge; 2018. p. 453–94.
- [41] Lin T-J, Nagpal M, VanDerHeide J, Ha SY, Newell GE. Instructional patterns for the teaching and learning of argumentative writing in high school English language arts classrooms. *Read Writ* 2020;33(10):2549–75. <https://doi.org/10.1007/S11145-020-10056-Y>.
- [42] Kruse, O., Anson, C.M. (2023). Writing and Thinking: what Changes with Digital Technologies?. In: Kruse, O., et al. *Digital writing technologies in higher education* Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_29.
- [43] Winne PH. Cognition and metacognition within self-regulated learning. In: Schunk DH, Greene JA, editors. *Handbook of self-regulation of learning and performance*. 2nd edition. Routledge/Taylor & Francis Group; 2018. p. 36–48. <https://doi.org/10.4324/9781315697048>.
- [44] Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3(2):77–101. <https://doi.org/10.1191/1478088706qp0630a>.
- [45] Lincoln YS, Guba EG. *Naturalistic inquiry*. Sage Publications; 1985.
- [46] Pearson PD, Gallagher MC. The instruction of reading comprehension. *Contemp Educ Psychol* 1983;8(3):317–44. [https://doi.org/10.1016/0361-476X\(83\)90019-X](https://doi.org/10.1016/0361-476X(83)90019-X).
- [47] Pintrich PR. Multiple goals, multiple pathways: the role of goal orientation in learning and achievement. *J Educ Psychol* 2000;92(3):544. <https://doi.org/10.1037/0022-0663.92.3.544>.
- [48] Ferretti RP, Lewis WE, Andrews-Weckerly S. Do goals affect the structure of students' argumentative writing strategies? *J Educ Psychol* 2009;101(3):577–89. <https://doi.org/10.1037/a0014702>.
- [49] Ray A, Fitzpatrick E. Practice-based professional development for self-regulated strategy development writing instruction with secondary teachers. *Teach Except Child* 2024;56(5):356–68. <https://doi.org/10.1177/0040059922112288>.
- [50] Travers NL, Sheckley BG. Changes in students' self-regulation based on different teaching methodologies. ERIC; 2000. ED445631, <https://files.eric.ed.gov/fulltext/ED445631.pdf>.
- [51] Zimmerman BJ. From cognitive modeling to self-regulation: a social cognitive career path. *Educ Psychol* 2013;48(3):135–47. <https://doi.org/10.1080/00461520.2013.794676>.
- [52] Hwang, Y., Lee, Y.M., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of new literacy skill using generative AI. *arXiv preprint arXiv:2311*. <https://doi.org/10.48550/arXiv.2311.05373>.
- [53] Raths LE, Harmin M, Simon SB. In: Merrill Charles E, editor. *Values and teaching: working with values in the classroom*; 1966.
- [54] Rest J, Narvaez D, Bebeau MJ, Thoma SJ. *Postconventional moral thinking: a neo-Kohlbergian approach*. Lawrence Erlbaum Associates; 1999.
- [55] Rodríguez Bermeo SD, Salazar ML, Parra Terán FF, Maldonado Calero JC, Albuja Centeno VJ. Impact of AI on virtual learning: self-sufficiency and academic confidence in university students. *Ciencia Latina* 2025;8(6):7414–29. https://doi.org/10.37811/cl_rcm.v8i6.15432.
- [56] Mehta V, Thomas V, Mathur A. AI-dependency in scientific writing. *Oral Oncol Rep* 2024;10. <https://doi.org/10.1016/j.oor.2024.100269>.
- [57] Floridi L, Cows J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Vayena E. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Mind Mach (Dordr)* 2018;28(4):689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- [58] Salo-Pöntinen H. AI ethics-critical reflections on embedding ethical frameworks in AI technology. In: *International Conference on Human-Computer Interaction*. Cham: Springer International Publishing; 2021. p. 311–29. https://doi.org/10.1007/978-3-030-77431-8_20.
- [59] Sharma V, Mishra N, Kukreja V, Alkhayyat A, Elngar AA. Framework for evaluating ethics in AI. In: *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*; 2023. p. 307–12. <https://doi.org/10.36571/ajsp661>.
- [60] Singh, A., Guan, Z., & Rieh, S.Y. (2025). Enhancing critical thinking in generative AI search with metacognitive prompts. *arXiv preprint arXiv:2505*. <https://doi.org/10.48550/arXiv.2505.24014>.
- [61] Stojković, N., & Wang, S. (2025). AI meta prompting as cognitive scaffolding in teaching academic writing. Available at SSRN 5295837.
- [62] Hamilton K, Piacentini M, Banister E. Ethics and care in qualitative research. In: Belk RW, Otnes C, editors. *Handbook of qualitative research methods in marketing*. Edward Elgar Publishing; 2024. p. 486–96. <https://doi.org/10.4337/9781035302727.00050>.
- [63] Kim J, Yu S, Lee SS, Detrick R. Students' prompt patterns and its effects in AI-assisted academic writing: focusing on students' level of AI literacy. *J Res Technol Educ* 2025;1–18. <https://doi.org/10.1080/15391523.2025.2456043>.
- [64] Chan STS, Lo NPK, Wong AMH. Enhancing university level English proficiency with generative AI: empirical insights into automated feedback and learning outcomes. *Contemp Educ Technol* 2024;16(4). <https://doi.org/10.30935/cedtech/15607>.
- [65] Tankelevitch L, Kewenig V, Simkute A, Scott AE, Sarkar A, Sellen A, Rintel S. The metacognitive demands and opportunities of generative AI. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, May; 2024. p. 1–24. <https://doi.org/10.1145/3613904.3642902>.