Using Learning Analytics to Measure AI-Critical Literacy: A Within-Subjects Study of the UnBlooms Framework

Tina R. Austin[1,*]

Abstract

This work-in-progress paper examines how learning analytics can operationalize AI-critical literacy by tracing students' metacognitive decision-making during AI-supported tasks. Using the UnBlooms Framework and its five-level Metacognitive Awareness Scale, the study analyzes reasoning artifacts from fifteen multidisciplinary undergraduate and graduate courses, including annotated drafts, AI interaction logs, structured reflections, and revision traces. Preliminary findings indicate movement from passive acceptance of AI outputs toward active verification, contextual reasoning, and self-critique. The paper introduces two analytic constructs–Decision Trails and Misconception Audits–as alternatives to AI-detection-based assessment, enabling instructors to diagnose reasoning processes rather than police tool use. The study contributes an analytics-informed approach for assessing learning in human-centered generative AI environments and outlines implications for equitable assessment design.

Keywords

Learning Analytics, Generative AI, Metacognition, AI-Critical Literacy, Assessment, UnBlooms, Epistemic Agency

1. Introduction

The rapid adoption of generative AI in educational contexts has intensified debates about academic integrity, cognitive offloading, and the future of assessment. Much of this discourse presumes that AI usage inherently diminishes thinking, positioning AI as a shortcut that bypasses learning. However, such claims often rely on assessment models that equate learning with the production of unaided final outputs–an assumption increasingly untenable in AI-mediated environments. As Bender et al. [3] demonstrated, large language models produce linguistically fluent text without any reference to meaning, making fluency an unreliable proxy for understanding.

Learning analytics offers an opportunity to shift the focus from what students produce to how students reason. Rather than attempting to detect or prohibit AI use, analytics can surface the decision-making processes students employ when interacting with machine-generated content. Christianson[4] argues that the detection arms race between students and faculty is unproductive and that educators should instead rethink assignments to surface genuine reasoning. This paper argues that AI-critical literacy should be assessed not through product authenticity but through evidence of epistemic agency: the ability to evaluate, contextualize, and refine AI outputs.

We report on a within-subjects study investigating whether structured instructional design, paired with analytic tracing of student reasoning artifacts, can reveal measurable growth in AI-critical literacy. The study is guided by the UnBlooms Framework, a problem-centered alternative to Bloom's Taxonomy that foregrounds metacognition, contextual judgment, and ethical reasoning in AI-augmented learning. [1, 2]

LA4GenAISLE'26: Learning Analytics for Human-Centered Generative AI-Enhanced Smart Learning Environments Workshop, April 28, 2026, Bergen, Norway

*Corresponding author.

tinaaustin@ucla.edu (T. R. Austin)

0009-0009-0719-7233 (T. R. Austin)

2. Conceptual Framework

2.1. The UnBlooms Framework

The UnBlooms Framework reframes learning progression away from linear cognitive hierarchies and toward recursive reasoning cycles. Rather than treating "creation" as an endpoint, UnBlooms positions AI-generated output as materia–raw material requiring evaluation, verification, and refinement.

The framework is operationalized through a CDIR model whose three dimensions–Context-Driven, Inclusive, and Reflective–are not merely analytic categories applied after the fact. They are engineered into the task design itself, operating sequentially across three instructional phases that make learner reasoning visible and measurable.

Before AI (Context-Driven). Prior to any AI interaction, the task design requires learners to surface the constraints that should govern their engagement: What disciplinary, ethical, or social constraints shape this task? When should AI not be used here? This phase anchors learners in their own epistemic context before AI-generated content can substitute for or distort that reasoning. It establishes a documented baseline against which subsequent AI outputs will be evaluated.

During AI (Inclusive Scaffolding). Required checkpoints interrupt the AI interaction at structured decision points rather than deferring reflection until after submission. Learners respond to embedded prompts: Why are you accepting this output? What claims require verification before you proceed? These checkpoints make otherwise invisible acceptance decisions traceable and generate the raw material for Decision Trail analysis. The Inclusive dimension ensures scaffolds do not privilege learners with prior domain expertise–prompts surface reasoning at whatever level of sophistication the learner currently operates.

After AI (Reflective). Structured reflection prompts are calibrated to the five levels of the Metacognitive Awareness Scale, requiring learners to justify decisions with specificity proportional to their demonstrated reasoning. A learner at Level 2 identifies a salient feature of the AI response; a learner at Level 4 explains the reasoning pattern they evaluated and why they revised or rejected the AI's approach. This calibration transforms reflection from a generic exercise into an analytically tractable record of epistemic agency.

The sequencing of these phases is critical: the Before AI stage establishes epistemic constraints, the During AI stage captures decision-making under those constraints, and the After AI stage elicits justification of those decisions. Together, this structure ensures that each phase generates analytically distinct data aligned with specific dimensions of epistemic agency.

This sequencing aligns with the Gradual Release of Responsibility model (Pearson and Gallagher [6] in which learners progressively assume greater cognitive responsibility. In AI-mediated contexts, this progression is inverted and intensified: as AI assumes generative functions, learners must assume greater evaluative and metacognitive responsibility. The CDIR structure operationalizes this shift by embedding deliberate intervention points that require learners to assert judgment before, during, and after AI interaction.

2.2. Metacognitive Awareness Scale

To support analytic measurement, UnBlooms includes a five-level Metacognitive Awareness Scale, ranging from surface-level error identification to advanced self-critique of one's own reasoning strategies. Zaphir et al.[5] demonstrate that critical thinking skills can only be mimicked by generative AI rather than genuinely performed, underscoring the need for frameworks that assess how students reason with AI rather than whether they used it. The scale classifies student artifacts based on observable indicators such as verification behavior, source triangulation, and revision rationales.

Table 1
The UnBlooms Five-Level Metacognitive Awareness Scale

Level	Behavior	Points
1	Recognizes differences between AI and own personal reasoning	1
2	Reflects on distinct features in AI output	2
3	Analyzes how AI's approach contrasts with human reasoning	3
4	Evaluates AI reasoning patterns and suggests improvements	4
5	Create or Resist: designs beyond AI limits or deliberately works without AI	5

Artifacts were coded by the author using a structured rubric aligned with the Metacognitive Awareness Scale. A subset (approximately 15–20%) was second-coded by an external educator for calibration; disagreements were resolved through discussion. Supplementary pre- and post-course survey data ($N = 529$ undergraduate respondents) were used for triangulation but are not the primary analytic focus of this paper. Table 1 illustrates the five levels.

3. Methods

3.1. Study Design

This study employs a within-subjects design, analyzing changes in student reasoning across multiple AI-mediated assignments within the same course. The dataset consists of instructional artifacts collected as part of normal coursework across fifteen multidisciplinary courses in life sciences, computational biology, social sciences, humanities, and professional programs.

3.2. Data Sources

Artifacts analyzed include:

- Annotated AI-assisted drafts
- Structured reflection prompts
- Revision histories
- Prompt-response interaction logs, where available

To support consistent analysis, artifacts were segmented into discrete decision points, such as acceptance, modification, or rejection of AI outputs, with each decision point treated as a unit within a Decision Trail sequence. This segmentation enables systematic comparison of reasoning behaviors across assignments and timepoints.

All data were de-identified prior to analysis and examined in aggregate.

Ethical note. This study analyzes existing instructional artifacts generated as part of routine educational practice. No experimental manipulation, intervention beyond standard pedagogy, or collection of sensitive personal data was conducted. Findings are reported at the aggregate level and presented as preliminary, exploratory results consistent with work-in-progress learning analytics research.

3.3. Analytic Constructs

Two analytic constructs were developed:

- Decision Trails: sequences of student choices indicating how AI outputs were evaluated, modified, or rejected.
- Misconception Audits: pattern analyses identifying recurring reasoning errors and verification failures across cohorts.

Extended example: CRISPR Decision Trail across scale levels. The following example traces one student's reasoning across three points in the same assignment, illustrating how the Metacognitive Awareness Scale captures qualitatively distinct forms of epistemic agency.

Level 4: Evaluative Critique. An AI-generated explanation of CRISPR off-target effects was initially accepted by the student. Subsequent annotations show the student cross-checking the claim against two external sources, flagging a terminology mismatch, and revising the explanation while documenting the rationale for rejecting the original AI phrasing. The student is operating on the AI's reasoning–evaluating its pattern and justifying a correction–but remains within the AI-assisted workflow.

Level 5 (Create): Task Redesign. After encountering repeated verification failures in AI-generated mechanistic biology explanations across multiple assignments, the same student redesigns their own AI interaction strategy. They develop a structured prompt template that requires the AI to cite specific molecular mechanisms and flag uncertainty, and a personal verification checklist calibrated to the epistemic standards of the domain. The student is no longer simply evaluating AI output; they are engineering the conditions of their own AI engagement. This was coded as Level 5 (Create): the learner has synthesized a new task structure that systematically constrains AI use based on disciplinary reasoning.

Level 5 (Resist): Deliberate Non-Use. In a subsequent module requiring mechanistic explanation of gene editing outcomes, the student reviews their Decision Trail from prior assignments and concludes that for this category of claim, the cognitive work of building understanding from primary literature is the learning goal–and that AI-generated explanations short-circuit exactly the inferential struggle that produces durable knowledge. The student documents this reasoning explicitly and completes the task without AI assistance. This was coded as Level 5 (Resist): the learner has exercised epistemic agency by deliberately opting out, with justification grounded in disciplinary and metacognitive judgment rather than avoidance.

Artifacts were coded using the Metacognitive Awareness Scale, enabling comparison of reasoning stages across timepoints. The CRISPR sequence illustrates that Level 5 is not simply "more" of Level 4: it requires a qualitative shift from operating on AI outputs to redesigning or opting out of the interaction itself. This pattern generalizes across domains: a writing student who builds a source-verification protocol before drafting, a programming student who implements an algorithm manually before using AI-generated code, and an ethics student who refuses AI assistance on a value-judgment task because reasoning through it unaided is the assignment all exhibit the same underlying structure of epistemic self-governance that defines Level 5.

This suggests that Level 5 is domain-independent but context-instantiated: the underlying cognitive move–exercising epistemic agency over whether and how to engage AI–is constant across disciplines, while the specific constraints, verification standards, and resistance rationales are supplied by the disciplinary context in which the learner is operating.

4. Findings

Preliminary analysis reveals consistent shifts in student reasoning behavior over successive AI-mediated tasks. Early artifacts frequently reflected uncritical acceptance of AI outputs, characterized by minimal verification and reliance on surface plausibility. Later submissions showed increased evidence of:

- external source triangulation;
- explicit justification for accepting or rejecting AI suggestions;
- recognition of contextual and ethical constraints;
- reflective critique of one's own reasoning process.

Misconception Audits identified common failure patterns, such as over-trust in fluent explanations and misapplication of domain-specific terminology. These insights enabled targeted instructional adjustments and informed subsequent task design.

5. Discussion

The findings suggest that AI-critical literacy can be meaningfully assessed by examining post-creation reasoning rather than final products. Learning analytics enables instructors to function as diagnosticians of reasoning, shifting assessment away from detection-based models[4] that disproportionately disadvantage under-resourced learners. By focusing on observable decision-making traces, this approach supports more equitable assessment practices while acknowledging the reality of AI-augmented knowledge work.

A key insight from this study is the distinction between analyzing reasoning and engineering the conditions under which reasoning becomes visible. The CDIR model's Before/During/After structure does both simultaneously: the instructional design choices that prompt learners before AI use, interrupt them during it, and elicit calibrated reflection afterward are the same choices that generate the analytic traces captured by Decision Trails and Misconception Audits. Assessment is not layered on top of instruction; it is embedded within it. This integration suggests the framework is transferable to other GenAI-enhanced learning contexts–including programming assistance, language learning, and vocational training–wherever the goal is to make the reasoning behind AI use visible and evaluable.

The inclusion of "Resist" as a valid Level 5 outcome reflects a broader theoretical commitment: epistemic agency includes knowing when not to use AI. For some learners and some tasks, the productive struggle of working without AI assistance is itself the learning goal. Both Create and Resist require the same depth of metacognitive awareness and disciplinary judgment, making them equally sophisticated achievements.

6. Limitations and Future Work

As a work-in-progress study, findings are exploratory and based on instructional data from a single institutional context with artifact coding conducted primarily by the author. Future work will compute formal inter-rater reliability statistics, expand the coding team, and pursue cross-institutional validation of the Metacognitive Awareness Scale.

A central goal for future iterations is the partial automation of Decision Trail and Misconception Audit extraction. The Before/During/After task structure produces structured data–timestamped checkpoints, embedded rationale prompts, revision logs–that are amenable to rule-based and machine-learning approaches. Automating even the flagging of Level 1 and Level 2 patterns, such as uncritical acceptance and absence of verification, would allow the framework to scale beyond single-instructor contexts. The CDIR task structure also appears adaptable to other AI tool types, including programming assistants and language learning applications, where the Before/During/After phases can be re-anchored to domain-specific epistemic constraints without altering the underlying analytic logic.

7. Conclusion

Generative AI does not eliminate thinking; it changes where thinking becomes visible. Learning analytics offers a path forward by capturing the reasoning students employ when navigating AI-generated content. By operationalizing AI-critical literacy through decision traces and misconception patterns, educators can design assessments that reflect how knowledge work is actually performed in AI-rich environments.

The UnBlooms Metacognitive Awareness Scale–including the Create/Resist distinction at Level 5–offers a practical, teachable, and analytically tractable framework for this work. It positions both

creative transcendence of AI and deliberate resistance to AI as equally valid expressions of epistemic maturity.

Declaration on Generative AI

During the preparation of this work, the author used Claude Sonnet 4.6 in order to check grammar and spelling. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content. All analytical claims, interpretations, and theoretical contributions are the original work of the author.

References

[1] T. R. Austin, The UnBlooms model: A problem-centered framework for learning design in the AI era, in: Oxford AI in Education Conference, 2025. https://doi.org/10.5281/zenodo.17298679

[2] T. Austin, J. Gulya, and M. Kassorla, Beyond the pyramid: Bloom's taxonomy meets generative AI, *Postdigital Science and Education*, manuscript submitted for publication, 2026.

[3] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, ACM, 2021, pp. 610–623. https://doi.org/10.1145/3442188.3445922

[4] J. S. Christianson, End the AI detection arms race, *Patterns* 5(10) (2024) 101058. https://doi.org/10.1016/j.patter.2024.101058

[5] L. Zaphir, J. M. Lodge, J. Lisec, D. McGrath, and H. Khosravi, How critically can an AI think? A framework for evaluating the quality of thinking of generative artificial intelligence, arXiv, 2024. https://doi.org/10.48550/arXiv.2406.14769

[6] P. D. Pearson and M. C. Gallagher, The instruction of reading comprehension, *Contemporary Educational Psychology* 8(3) (1983) 317–344.