

# Cognitive Integrity in Practice

A preview of Banyan's approach, with sample rubric rows and our research grounding

---

**What this document is.** A preview of how Banyan thinks about AI cognitive integrity in K-12 instruction. It includes excerpts from the Cognitive Integrity Rubric (Grade 8), highlights from the Assessment Integrity Companion, and a research-grounding section that names the specific work our position draws on, including studies that point in different directions.

**What this document is not.** The full deliverable. The complete Rubric (48 task classifications, four subjects) and the full Companion (8 sections, including detection signals, conferencing protocols, grading frameworks, family communication templates, and cross-classroom coherence guidance) are part of what districts receive through a Banyan Readiness Scan, Audit, or Blueprint engagement. This preview is what we make available before that engagement — enough for districts and their teams to evaluate whether the approach is worth pursuing.

**Why this is public.** The underlying research on AI and learning is in early days. Banyan publishes its position openly so districts can examine it, push back on the parts they disagree with, and decide whether this is an approach they want to engage with. The rubric is a starting point that districts are expected to adapt.

## 1. Our position

The Cognitive Integrity Rubric is grounded in two things: Banyan Global Learning's seventeen years of practitioner experience designing and delivering K-12 instruction since 2009, and the cognitive science literature on learning, productive struggle, and skill development.

Banyan has designed creative units of instruction across ELA, math, science, social studies, and world languages, delivered through our own teaching methodology to students in classrooms around the world. The question at the center of the Rubric

— what cognitive work must students still perform for learning to occur — predates the LLM era. Banyan teachers have been asking it about every form of cognitive scaffolding since 2009; what's new is applying the same question to a more capable tool.

The Rubric formalizes that question for an AI era: if we know which cognitive operations build a particular skill, we know which operations cannot be delegated to AI without undermining the skill itself. The Companion that pairs with the Rubric addresses how to make that classification operationally enforceable in actual classrooms.

## 2. The classification system

The Rubric sorts every common K-12 task type into one of three categories — or a fourth transitional category that names a developmental progression.

**HUMAN-ESSENTIAL** — This task IS the learning. The cognitive work is inseparable from the skill development. If AI performs it, the learning does not happen. Example: argumentative essay drafting, hypothesis formation, geometric proof.

**AI-AMPLIFIED** — Human work comes first; AI enhances afterward. The human cognitive operation is the anchor; AI adds efficiency, perspective, or feedback after the thinking is done. Example: grammar editing on a student-written draft, computational practice after conceptual mastery, citation formatting.

**AI-LITERACY** — A new skill students must explicitly learn about AI itself: capabilities, limitations, biases, evaluation criteria, and AI's role in their discipline. Example: evaluating AI-generated historical narratives, prompt engineering for research, identifying bias in AI output.

**HUMAN-ESSENTIAL (→ AI-AMPLIFIED)** — A transitional classification. The task is HUMAN-ESSENTIAL during the period when the underlying skill is being built, and moves to AI-AMPLIFIED once the student has demonstrated competency. Example: lab report writing, summarization of complex texts, revision and self-editing. The Companion discusses how to operationalize the transition through teacher attestation of mastery.

Why the system has three categories (plus a transition) and not four or seven. We considered a four-tier model with a case-by-case middle category and rejected it. The value of the Rubric is that it produces a defensible classification for every

common task type. Adding ambivalence categories preserves teacher uncertainty without resolving it; districts that need decisions cannot use a tool that defers them.

### 3. Sample rubric rows

Five rows drawn from the published Grade 8 Rubric. The complete published Rubric contains 48 task classifications across ELA, math, science, and social studies. K-5 and 9-12 grade-band versions are in development.

#### Sample 1 — ELA · HUMAN-ESSENTIAL

##### **Argumentative Essay Drafting** · *CCSS.ELA-LITERACY.W.8.1* **HUMAN-ESSENTIAL**

**Rationale.** Formulating claims and marshaling evidence IS the thinking. The struggle to construct and defend a position builds critical reasoning. AI-assisted drafting bypasses the cognitive operation that the standard exists to develop.

**Implementation.** Students should draft without AI assistance. AI may review final drafts for feedback, but argument construction must remain human-driven. Use process artifacts (Google Docs version history) and in-class anchors to make the classification enforceable.

#### Sample 2 — ELA · AI-AMPLIFIED

##### **Grammar & Mechanics Editing** · *CCSS.ELA-LITERACY.L.8.2* **AI-AMPLIFIED**

**Rationale.** Mechanical editing is not the learning target. After students produce their own draft, AI can assist with grammar, punctuation, and mechanics checking. Practitioner evidence: students who self-edit first and then use a tool to catch missed errors develop editing skill faster than students who self-edit twice.

**Implementation.** Students write and self-edit first. Then use AI tools to catch missed errors and polish conventions. Require a one-paragraph disclosure: which tool was used, what was kept, what was rejected.

#### Sample 3 — Math · HUMAN-ESSENTIAL

## Novel Problem-Solving with Expressions/Equations ·

*CCSS.Math.Content.8.EE.1-8*

### HUMAN-ESSENTIAL

**Rationale.** Mathematical reasoning develops through productive struggle. Students must grapple with unfamiliar problem structures to build conceptual understanding. AI-generated solutions short-circuit the struggle that builds the skill.

**Implementation.** Students should attempt novel problems independently first. Do not use AI to generate solutions. In-class anchoring of the first attempt produces the process artifact that subsequent work has to be consistent with.

## Sample 4 — Science · AI-LITERACY

### Evaluating AI-Generated Scientific Claims · *MS-LS1, MS-ESS2*

#### AI-LITERACY

**Rationale.** Students need to learn to critically evaluate AI-generated content in scientific contexts. AI tools produce plausible-sounding scientific claims that may be inaccurate or unsupported. The skill of evaluating those claims is itself a new science literacy.

**Implementation.** Provide students with AI-generated scientific claims and ask them to evaluate validity, sourcing, and alignment with established evidence. Teach them to identify confident-sounding fabrication and to ask for citations.

## Sample 5 — Social Studies · AI-LITERACY

### Evaluating AI-Generated Historical Narratives · *C3 Framework*

*D2.His.1-5*

#### AI-LITERACY

**Rationale.** AI tools generate historical narratives that may oversimplify, omit perspectives, or reflect training-data biases. Students need to critically evaluate AI-generated historical accounts as a new literacy skill in the discipline.

**Implementation.** Provide AI-generated historical narratives and have students compare them against primary sources, identify omissions and biases, and reconstruct the underlying perspectives the AI account elides.

*The five rows above represent three of the four classifications (HUMAN-ESSENTIAL, AI-AMPLIFIED, AI-LITERACY). The fourth — HUMAN-ESSENTIAL (→ AI-AMPLIFIED) — is used in the Rubric for tasks like Summarization of Complex Texts, Lab Report Writing, and Revision and Self-Editing, where the classification shifts as the underlying skill is established. The Companion describes how to operationalize the transition.*

## 4. Highlights from the Assessment Integrity Companion

The Rubric tells teachers which tasks belong in which classification. The Companion addresses how to make a HUMAN-ESSENTIAL classification enforceable when students have AI in every device they touch. Four highlights from the Companion that illustrate its operational stance:

### **Highlight 1 — On AI-detection software**

The Companion takes an explicit and unfashionable position on Turnitin AI, GPTZero, Originality.ai, Copyleaks, and similar tools: they are not reliable enough to anchor an academic integrity case. As of 2026, peer-reviewed studies and vendors' own published accuracy data show false-positive rates in the 5 to 30 percent range, with substantially worse false-positive rates on writing from English-language learners, neurodivergent students, and students using accessibility supports. A teacher who relies on detection software as primary evidence will, on average, falsely accuse roughly one student in every twenty. The Companion treats detection software as a confirming signal only — never as the primary case.

### **Highlight 2 — The four-mechanism enforcement stack**

Instead of relying on detection, the Companion describes four classroom mechanisms that produce evidence of authorship without algorithmic guessing:

- **Process artifacts** — Google Docs version history, in-class handwritten outlines, iterative draft submissions. Produces evidence during the work, separate from the final product.
- **Oral defense and conferencing** — 3-to-5-minute conversations during work time on a random subset of assignments. A student who wrote the essay can describe their thesis selection, their stuck moments, and how they'd extend the argument; a student who outsourced the work cannot.
- **In-class anchors** — Doing the high-stakes cognitive work during class time, with phones in caddies. Take-home work becomes preparation and revision; the assessment of cognitive work happens where the teacher can see it.

- **Authentic context constraints** — Prompts written so specifically to the local class context — references to specific discussions, dates, local issues — that generic AI output is identifiable on its face.

Most teachers run two or three of the four; few run all four on every assignment. Choice depends on subject and classroom culture.

### **Highlight 3 — Two-axis grading for AI-AMPLIFIED tasks**

When an assignment is AI-AMPLIFIED by design, the rubric evaluates two distinct things: the human contribution (the original thinking the student did before AI was introduced, weighted 60–70%) and appropriate AI use (was AI introduced at the right phase, were suggestions evaluated rather than just accepted, was the source disclosed, weighted 30–40%). A required one-paragraph disclosure at the end of every AI-AMPLIFIED submission becomes part of the grade.

### **Highlight 4 — Honest engagement with contestable classifications**

The Companion includes a section titled "The four most contestable classifications" that names rubric rows specialists will reasonably push back on. Each is presented with the strongest version of the counter-argument:

- **Grammar editing classified AI-AMPLIFIED.** Pushback: for English-language learners, syntactic editing IS the learning. The classification should revert to HUMAN-ESSENTIAL during the L2 acquisition phase. The Companion agrees and provides a local-adaptation path.
- **Citation formatting classified AI-AMPLIFIED.** Pushback from library specialists: citation construction is research-integrity training, not mechanical formatting. The Companion agrees that citation should remain HUMAN-ESSENTIAL for the first 2–3 research projects in a grade band, then transition.
- **Vocabulary acquisition in context classified HUMAN-ESSENTIAL.** Pushback from reading specialists: explicit vocabulary instruction produces stronger retention than contextual inference alone for below-grade-level readers. The Companion acknowledges the literature supports both approaches and recommends using them together.
- **Computational practice classified AI-AMPLIFIED.** Pushback from math specialists: the qualifier "once conceptual mastery is achieved" carries weight. AI-assisted computation for students who have not yet built fluency short-circuits the next grade band. The Companion recommends teacher attestation of mastery per student before AI use is permitted on these tasks.

This is the section of the Companion that most differentiates Banyan's approach from a one-size-fits-all rubric. The classifications are starting positions; specialists who know their domain better are expected to surface where local adaptation is warranted, and the Companion provides the structure for that conversation.

## 5. Research grounding

The Rubric's pedagogical position is informed by three bodies of research: the established cognitive science on learning and skill development, the recent (2023–2025) research on AI's effects on cognition and learning, and Banyan's seventeen years of practitioner observation. We address each in turn, and we name where these bodies of work disagree.

### 5a. Established cognitive science

The older literature is more robust and more settled than the contemporary AI-and-learning research. Our pedagogical position draws primarily from this foundation.

#### **Robert Bjork — Desirable difficulties**

Bjork's body of work at UCLA over three decades is the foundation for our position that productive struggle in foundational cognitive operations builds more durable skill acquisition than frictionless delivery. The principle that *learning that feels harder while it is happening produces stronger long-term retention* is one of the most replicated findings in the cognitive science of learning.

*Representative source: Bjork, E. L. & Bjork, R. A. (2011). "Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning." In Gernsbacher, Pew, Hough, and Pomerantz, eds., Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society. Worth Publishers.*

#### **John Sweller — Cognitive Load Theory**

The framework for understanding when cognitive offloading helps learning (lifting extraneous load that doesn't build the target skill) and when it harms learning (lifting germane load that IS the target skill). This is the conceptual basis for the AI-AMPLIFIED vs. HUMAN-ESSENTIAL distinction in the Rubric.

*Representative source: Sweller, J. (1988). "Cognitive Load During Problem Solving: Effects on Learning." Cognitive Science, 12(2). Subsequent work over thirty years has extended and complicated the framework.*

### **Daniel Willingham — Classroom-applied cognitive science**

Willingham's body of work translates cognitive science research into actionable guidance for K-12 teachers. Particularly relevant to the Rubric: his work on memory, attention, feedback timing, and the conditions under which understanding is built rather than merely demonstrated.

*Representative source: Willingham, D. T. (2009; 2nd edition 2021). Why Don't Students Like School? A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom. Jossey-Bass.*

### **Roediger & Karpicke — Retrieval practice**

Foundational research on the testing effect — that retrieving information from memory produces stronger learning than re-studying the same material. Relevant to the Rubric because AI assistance during cognitive work often replaces the retrieval moment that builds durable memory.

*Representative source: Roediger, H. L., III & Karpicke, J. D. (2006). "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." Psychological Science, 17(3).*

### **Risko & Gilbert — Cognitive offloading (pre-LLM foundation)**

The foundational paper on what happens when we offload cognitive work to external tools. Published before ChatGPT but the framework applies directly to LLM use. Distinguishes when offloading helps performance versus when it undermines the development of internal capacity.

*Representative source: Risko, E. F. & Gilbert, S. J. (2016). "Cognitive Offloading." Trends in Cognitive Sciences, 20(9).*

## **5b. Recent AI-and-learning research**

**This is the contested literature.** The 2023–2025 research on AI's effects on cognition and learning is moving fast, and the findings frequently contradict each other. Below are studies on multiple sides of the central question — does AI assistance help or harm cognitive development. Our classifications lean toward preserving human cognitive work where the evidence is contested. That's a precautionary stance, not neutrality, and it's worth naming as such. We've tried to land on calls that hold up reasonably well across the disagreements in the literature.

**A note on citation status.** Where studies below are preprints or working papers, we say so. The contemporary research is moving faster than peer review, and several important findings have not yet completed full review.

### **Studies suggesting AI use can harm cognition or learning**

**Kosmyna et al., MIT Media Lab (2025).** "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task." Used EEG to measure neural activity during essay writing. Found significantly reduced neural connectivity (alpha and theta bands) in participants who used LLMs compared with participants who used Google or no tools. Introduced the framing of "cognitive debt" — the idea that AI assistance during cognitive work creates a deficit that accumulates over time. *Preprint; peer review in progress.*

**Lee, Sarkar et al., Microsoft Research & Carnegie Mellon (2025).** "The Impact of Generative AI on Critical Thinking." Survey of 319 knowledge workers found that higher confidence in AI was associated with less critical engagement with AI outputs. Workers who trusted AI tended to perform less verification and to engage in less deep evaluation of the material AI produced. The implication for student learning: students who develop AI trust without first developing critical evaluation skills may be at greatest risk of skill erosion.

**Sparrow, Liu & Wegner (2011).** "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science*. Pre-LLM but conceptually parallel: when participants believed information would remain available externally, they remembered the information less well and remembered where to find it instead. The same dynamic has been observed in early studies of LLM users.

### **Studies suggesting AI use can help learning when used thoughtfully**

**Mollick & Mollick, Wharton (2024).** "Assigning AI: Seven Approaches for Students." Identifies seven pedagogical patterns for productive AI use in education — AI as tutor, coach, mentor, teammate, tool, simulator, and student. An influential framework drawing on extensive practitioner observation rather than empirical study. Argues that the question is not whether to integrate AI but how to do so in ways that build rather than substitute for thinking. *Working paper.*

**Kestin et al., Harvard (2024).** Preprint reporting that AI-tutored students in an undergraduate physics course outperformed students in active-learning lectures. The study was small and limited to one subject and one institution, and the results have been contested, but it represents the most-cited recent finding suggesting AI tutoring can outperform high-quality conventional instruction in specific contexts.

**Khan Academy / Khanmigo research (2024-2025).** Multiple studies of Khan Academy's AI tutor (Khanmigo) in school deployments. Results are mixed: some implementations show learning gains, others show no significant effect, and the question of which student populations benefit most is unresolved. Khan Academy itself has framed Khanmigo as a possible approach to Bloom's "2 sigma problem" (the finding that one-on-one tutoring produces gains two standard deviations above conventional instruction).

### **The contested middle**

Even within the research that finds positive AI effects, the studies disagree on:

- Subject area — AI tutoring shows different effects in math versus writing versus science.
- Student ability — high-performing students may benefit from AI in ways struggling students do not, or vice versa, depending on the study.
- When in skill development — AI use after foundational skills are built may be different from AI use during foundational skill-building.
- Type of AI use — AI suggestion (student decides) is not the same as AI completion (AI writes, student edits) is not the same as AI tutoring (AI asks questions and guides).

These dimensions are largely unresolved in the published literature as of 2026. The Rubric's classifications attempt to land defensible positions across this contested terrain. They will need revision as the research accumulates.

## **5c. Practitioner observation**

Where the academic literature suggests one classification and Banyan's classroom evidence suggests another, we have favored the classroom evidence in the published Rubric. Two examples already mentioned in Section 4: grammar editing classified AI-AMPLIFIED on practitioner evidence (faster skill acquisition through tool-assisted feedback) even though a strict reading of cognitive load theory would suggest HUMAN-ESSENTIAL; and lab data collection classified HUMAN-ESSENTIAL on practitioner evidence even though some recent AI tools claim assistive capability.

Practitioner evidence is not a substitute for empirical research. It is, however, the form of evidence that K-12 teachers actually generate and that the Rubric must be defensible against. Banyan's seventeen years of unit design and delivery is the backbone against which we have checked every academic finding before publishing a classification.

## 6. What this Rubric is not

This is a starting position, not the final word. The research will continue to evolve and so will our classifications. The Companion includes a section on the four most contestable rows; reasonable specialists who read the same evidence may reach different calls.

The contemporary research on AI and learning is not settled. Our classifications lean toward preserving human cognitive work where the evidence is contested — a precautionary stance we want to acknowledge rather than disguise as neutrality. Established cognitive science, the contested newer literature, and seventeen years of practitioner observation all inform the calls. Where local evidence justifies a different call, the rubric expects districts to make it.

Districts that adopt the Rubric are expected to adapt it. The Companion provides the structure for that through the quarterly department alignment check and the audit checklist. A row that's wrong for your context should be changed.

Future versions are published as our understanding of the field changes. Districts engaged with Banyan receive updates as part of ongoing partnership; the public Rubric is updated on [banyangloballearning.com](http://banyangloballearning.com).

## 7. How the Rubric evolves

Three mechanisms shape future versions of the Rubric:

- **New empirical research.** As new studies on AI and learning are published — whether they support or challenge current classifications — Banyan reviews them and updates the Rubric where the evidence warrants. The methodology for review is the same one applied to the current version: established cognitive science as the foundation, recent AI research as informative but not authoritative, practitioner observation as the backbone.
- **Implementation reports from districts.** Districts and teachers using the Rubric in classroom settings produce evidence about which classifications work in practice and which do not. Banyan reads and incorporates this feedback into subsequent versions. Implementation reports are particularly valuable when they identify rows that look defensible in theory but fail in classroom conditions.
- **Specialist review.** Subject-matter specialists — ELL teachers, math coaches, library media specialists, reading specialists — surface where the

Rubric's general classifications need local adaptation. The Companion's contestable-classifications section is built on this kind of review and will expand as more specialists weigh in.

Version dating. Every published version of the Rubric is dated and versioned. Implementations in the field are encouraged to record which version is in use. Banyan does not retroactively change classifications without explicit version notes.

## 8. Note on AI assistance in production

Banyan worked with **Claude** (Anthropic's language model, accessed through the Cowork product) to draft, refine, and standards-align portions of the Rubric and the Companion. The framework, the four-classification system, the pedagogical position, the practitioner-grounded calls, and the editorial voice are Banyan's. Claude assisted with synthesis of source material, draft language, and verification of standards citations. Every cell in every row of the published Rubric was reviewed by a Banyan curriculum lead before publication.

We name this because the subject of the Rubric is AI assistance, and being transparent about the inputs into the work seemed like the right starting point.

## If you'd like to discuss this

Banyan Global Learning is a 17-year-old K-12 services company. We help districts make AI decisions deliberately — what to allow, what to teach, what to buy, and how to keep it consistent across schools. The Cognitive Integrity Rubric and Companion are part of our AI Governance service line.

To take our free 10-minute AI Readiness Snapshot — a self-administered survey that produces a 2-page maturity report for your district — visit [banyangloballearning.com/ai-governance](https://banyangloballearning.com/ai-governance). To talk with us about a Readiness Scan, Audit, Blueprint, or Retainer engagement, email [travis@banyangloballearning.com](mailto:travis@banyangloballearning.com).

Feedback, redlines, or implementation reports from districts in the field are read and incorporated into subsequent versions of the Rubric and the Companion. The Cognitive Integrity Kit improves through use.

*Approach Preview v1 — compiled June 2, 2026. The complete Cognitive Integrity Rubric (Volume I) and Assessment Integrity Companion (Volume II) are delivered as part of Banyan Readiness Scan, Audit, or Blueprint engagements. Volumes III (Equity Companion) and IV (Lesson Plan Exemplar) are in development. © 2026 Banyan Global Learning.*