International Journal of Educational Development
ISSN Print: 3104-4263
ISSN Online: 3104-4271
CD ISSN: 3104-428X
https://doi.org/10.63313/IJED.9119

Making AI-Assisted Academic Writing Visible: Mapping Assessment Evidence to Learning Outcomes in Undergraduate Academic Writing

Qinyu E
College of Publishing, University of Shanghai for Science and Technology, Shanghai 200093, China
Email: eqinyu0327@gmail.com

How to cite this paper: E, Q. (2026). Making AI-assisted academic writing visible: Mapping assessment evidence to learning outcomes in undergraduate academic writing. International Journal of Educational Development, 4(3), 78–88. ISSN Print: 3104-4263, ISSN Online: 3104-4271.
https://doi.org/10.63313/IJED.9119
Published: 2026-10-30

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

Abstract

Generative artificial intelligence (GenAI) complicates the assessment of learning in undergraduate academic writing courses: a polished paper does not necessarily reveal what students understand or can do independently. Prior work has examined writing portfolios, student–AI interactions, and broader assessment approaches, but instructors still need a practical way to connect evidence from writing activities to specific learning outcomes and judge what that evidence supports. This conceptual paper proposes an assessment-evidence-to-learning-outcome mapping framework for undergraduate academic writing instruction. Drawing on constructive alignment, assessment validity, portfolio scholarship, DRIVE, and the 4Ps model, it defines four adaptable course-level outcomes: problem formulation, argumentative reasoning, critical verification, and ethical responsibility. It maps product, process, performance, and practice evidence to these outcomes and illustrates how complementary sources can inform bounded assessment judgments through a constructed case and an illustrative rubric. The framework is a context-specific design proposal, not a validated instrument; its value and feasibility require empirical testing.

Keywords

Generative Artificial Intelligence; Undergraduate Academic Writing Instruction; Learning Outcomes; Assessment Evidence; Assessment Validity

1. Introduction

Generative artificial intelligence (GenAI) has changed how undergraduate academic writing is taught, produced, and assessed. Students may use AI to explore topics, develop arguments, summarize literature, and revise texts. While these tools can support learning, a polished paper may reflect students' own understanding, collaboration with AI, or substantial delegation of intellectual work. The final text alone cannot reliably distinguish among these possibilities.

Although final written products remain important, their quality provides only partial

evidence of student learning. Xia et al., reviewing 32 empirical studies, highlighted self-regulated learning, responsible learning, and academic integrity as key considerations in higher-education assessment.[1] Meng et al., synthesizing 25 studies of AI-assisted academic writing, found more consistent benefits for psychological and process-related outcomes than for textual quality, with mixed findings on higher-order skills such as argumentation and discourse organization.[2] These findings caution against equating improved writing with improved learning.

Existing approaches provide important foundations: writing portfolios document development across drafts;[3] DRIVE examines student–GenAI interactions;[4] the 4Ps framework distinguishes product, process, performance, and practice as assessment evidence;[5] and Nikolić and Basta Nikolić propose an AI-resilient assessment framework[6]. However, a practical challenge remains: how can writing artifacts and activities be linked to specific learning outcomes, and how can evidence from different sources support defensible judgments? Prompt logs, revision histories, and AI-use declarations provide useful information but cannot independently establish students' understanding or improved reasoning. The key challenge is therefore to interpret and corroborate evidence in relation to intended learning outcomes. Accordingly, the present paper addresses two research questions:

RQ1. How can evidence generated during AI-assisted undergraduate academic writing instruction be mapped onto specific learning outcomes?

RQ2. How can process evidence, final written work, students' explanations, and authentic writing activities be combined to support transparent and defensible judgments about student learning?

To address these questions, this paper proposes an assessment-evidence-to-learning-outcome mapping framework, identifies four course-relevant learning outcomes, and outlines principles for corroborating evidence. A constructed case illustrates its application in a staged writing course. The paper offers a context-specific conceptual and pedagogical design, not a validated assessment instrument or evidence of improved learning outcomes.

2. Conceptual Background and Literature Positioning

2.1. Constructive alignment and assessment validity

Constructive alignment links intended learning outcomes with teaching activities and assessment tasks.[7] In writing instruction, tasks should elicit evidence of the capabilities being taught—such as formulating a focused problem, reasoning with evidence, verifying information, and using sources responsibly—rather than reward linguistic polish alone. Alignment does not, however, guarantee validity. Messick defines validity in terms of the interpretations and decisions made from assessment evidence.[8] The key question in AI-assisted writing is therefore whether available evidence warrants conclusions about student learning. The framework links each outcome to relevant evidence, compares sources where feasible, and limits

judgments to what the evidence supports.

2.2. Writing portfolios and process documentation

Writing portfolios make development visible by preserving drafts, feedback, revisions, reflections, and final submissions. Lam's review highlights their potential to support reflection and metacognition, while noting a gap between portfolio scholarship and classroom implementation.[3] In AI-assisted writing, however, multiple drafts or reflections do not necessarily demonstrate understanding: repeated AI rewriting may produce visible changes without showing that students can explain or justify them. The present framework therefore treats portfolio artifacts as selective evidence, not proof in themselves. It prioritizes consequential decisions, such as narrowing a topic, rejecting an unsupported claim, or restructuring an argument, and does not treat portfolio completeness as a proxy for competence.

2.3. Assessing learning through AI interactions: the contribution and limits of DRIVE

Oliveira et al. developed DRIVE to examine learning through student–GenAI interactions.[4] It focuses on Directive Reasoning Interaction (DRI), or how students direct and reason through interactions, and Visible Expertise (VE), or how disciplinary knowledge appears in those interactions. Analyzing 1,450 interactions linked to 70 graded essays in STEM writing-intensive courses, they found a strong positive correlation between interaction quality and essay scores.

DRIVE makes interaction records a potential source of learning evidence, but its reported correlation does not establish causation or guarantee transfer across disciplines and tasks. Since DRIVE already links interaction patterns to learning objectives, the present paper does not claim outcome alignment as new. Instead, it places AI interactions alongside other course evidence—topic notes, source checks, outlines, revisions, final texts, student explanations, and authentic writing tasks—to clarify what can be inferred from their combination.

2.4. The 4Ps model and AI-resilient assessment

The 4Ps framework distinguishes product, process, performance, and practice as different proxies for learning, each with its own strengths and limitations.[5] A final paper, a record of its development, an oral explanation, and work in an authentic activity may reveal different aspects of capability, but none provides a complete account on its own. Nikolić and Basta Nikolić propose a complementary four-pillar approach to AI-resilient assessment: process documentation, oral defense, authentic task design, and transparent AI-use policies aligned with learning outcomes.[6] Their conceptual paper also offers tools for future evaluation without claiming empirical validation.

2.5. Comparative positioning of existing frameworks

The comparison below summarizes how the main approaches inform the proposed design. It describes their roles in this paper, not their relative effectiveness.

Table 1. Relationship Between Existing Approaches and the Proposed Framework

Approach	Main contribution	Role in the present framework
Constructive alignment (Biggs & Tang, 2011)	Aligns learning outcomes, teaching activities, and assessment	Connects course goals with instruction and assessment; validity theory adds scrutiny of the inferences drawn from evidence
Writing portfolios (Lam, 2017)	Documents drafts, feedback, revision, and reflection	Informs selective collection of consequential artifacts and interpretation against explicit outcomes
DRIVE (Oliveira et al., 2025)	Examines student–GenAI interactions and relates interaction patterns to learning objectives	Treats interaction records as one evidence source within a broader course-level map
4Ps (Fawns et al., 2026)	Distinguishes product, process, performance, and practice evidence	Organizes evidence types and highlights their limits as proxies for learning
Four-pillar AI-resilient assessment (Nikolić & Basta Nikolić, 2026)	Combines process documentation, oral defense, authentic tasks, and transparent AI-use policies	Informs course design while the present framework specifies writing outcomes and links them to evidence

2.6. Deriving the four learning outcomes

The four outcomes are theory-informed course-design dimensions, not a universal or empirically validated taxonomy of academic writing competence. They represent distinct but related demands of academic writing and risks associated with AI-assisted work.

- **Problem formulation** concerns establishing a focused, meaningful question or line of inquiry. Constructive alignment emphasizes observable outcomes, while research-writing guidance links a research problem and question to a defensible argument.[7,9] The ACRL Framework also presents research as iterative inquiry in which questions are refined over time.[10]
- **Argumentative reasoning** concerns connecting claims, reasons, and evidence into a coherent argument. Booth et al. describe research writing as a process of developing problems, claims, reasons, evidence, and responses to readers' questions.[9] This outcome differs from problem formulation by focusing not on the inquiry's scope, but on how claims are developed and supported.
- **Critical verification** concerns evaluating information credibility and relevance, checking AI-generated claims, and ensuring that cited sources support the statements attributed to them. Information-literacy principles emphasize contextual evaluation of authority, strategic searching, and critical engagement with information, including AI-generated content.[10] Verification asks whether information is credible and accurately represented; argumentative reasoning asks whether it supports the argument.
- **Ethical responsibility** concerns transparent and appropriate AI use, accurate attribution, compliance with course requirements, and accountability for submitted work. Research on assessment redesign highlights responsible learning and academic integrity,[1] while the ACRL Framework emphasizes ethical participation in scholarly communication.[10] An AI-use declaration can support transparency, but cannot by itself establish complete disclosure or

responsible conduct.

3. Framework Development Approach

This paper uses a focused conceptual synthesis and pedagogical design approach, not a systematic review or empirical evaluation. Sources were selected to connect constructive alignment and assessment validity,[7,8] writing portfolios,[3] AI-assisted assessment,[4,6] and research writing, information literacy, academic integrity, and AI-assisted writing outcomes.[1,2,9,10]

The framework was developed in five steps: defining the assessment problem; comparing the roles of existing approaches; selecting four connected outcomes—problem formulation, argumentative reasoning, critical verification, and ethical responsibility; mapping these outcomes to product, process, performance, and practice evidence; and illustrating corroboration through a rubric and a constructed case. Four principles guide the design: outcome relevance, evidence complementarity, corroboration across sources where feasible, and proportionality in workload, accessibility, and privacy. Because the source selection was not exhaustive and no formal screening, independent coding, or empirical validation was conducted, the mapping and rubric remain proposals for local adaptation and testing.

Table 2. Illustrative alignment of course goals, tasks, outcomes, and evidence

Course goal	Example task or activity	Outcome	Relevant evidence and criteria
Develop a focused inquiry	Topic-narrowing exercise and research paper	Problem formulation	Initial and revised question; decision rationale; final paper. Assess focus, relevance, scope, and justification.
Build an evidence-based argument	Outline workshop, peer review, and revision	Argumentative reasoning	Outline versions, revision rationale, final paper, explanation of a claim. Assess coherence and links among question, claims, evidence, and conclusions.
Evaluate information accurately	Source-evaluation and claim–citation checks	Critical verification	Verification record, retrievable sources, relevant passages, explanation. Assess authenticity, relevance, accurate representation, and correction of unsupported claims.
Use AI and sources responsibly	AI-policy briefing, citation practice, disclosure	Ethical responsibility	AI-use statement, selected records, citations, explanation of responsibility. Assess transparency, policy compliance, attribution, and accountability.

This table is illustrative, not a prescribed curriculum. Add or substitute outcomes when local goals are not represented, and tell students what to retain, why it matters, how it will be assessed, and how records will be handled.

4. An Assessment-Evidence-to-Learning-Outcome Mapping Framework

4.1. Defining the Learning Outcomes

The framework uses four connected outcomes relevant to undergraduate academic writing instruction:

- **Problem formulation** is the ability to establish a focused and meaningful writing problem, question, or thesis, define its scope, and explain its relevance. Students may use GenAI to explore topics but remain responsible for judging focus and feasibility.

- **Argumentative reasoning** is the ability to connect claims, evidence, and reasoning coherently, and to explain how the paper's question, structure, evidence, and conclusions relate.
- **Critical verification** is the ability to evaluate information quality, verify source authenticity, check whether sources support attributed claims, and correct factual or logical errors. It involves more than confirming that a reference exists.
- **Ethical responsibility** is the ability to use AI transparently and appropriately, follow citation and authorship requirements, comply with course policies, and take responsibility for the submitted work. Disclosure is relevant but insufficient on its own to establish responsible use.

These categories are adaptable course-level dimensions, not fixed personal traits or an exhaustive taxonomy.

4.2. Mapping Evidence to Learning Outcomes

The framework uses the 4Ps model to organize, rather than replace, existing evidence categories. Product evidence appears in the submitted text; process evidence records selected stages and decisions; performance evidence comes from students' explanations or demonstrations; and practice evidence comes from authentic course activities, such as topic-narrowing, source evaluation, peer review, or guided revision.

Table 3. Mapping evidence to learning outcomes

Outcome	Process evidence	Product evidence	Performance evidence	Practice evidence
Problem formulation	Topic notes, selected prompts, changes to question or thesis, decision rationale	Final question/thesis and scope	Explanation of topic choice and scope	Completed topic-narrowing task with a rationale for the final question
Argumentative reasoning	Outline versions, structural changes, revision rationale	Organization and links between claims and evidence	Explanation of a claim or structural decision	Annotated outline or peer-review response showing how feedback was used to test the argument
Critical verification	Source-checking notes, fact-checking record, correction log	Retrievable references that support associated claims	Explanation or demonstration of a verification decision	Completed claim–citation check with a recorded decision about source support
Ethical responsibility	Relevant AI-use records, disclosure preparation, reflection	AI-use statement, attribution, compliant citations	Explanation of AI-use boundaries and responsibility	Completed disclosure or attribution task applying the course AI policy to a specific use

The matrix identifies potentially relevant evidence, not proof. A prompt log may show exploration without a focused question; an outline may change without stronger reasoning; and a verification record may document a check without showing that a source supports the claim. Classroom performance may also fail to generalize across contexts. Instructors should state what each source supports, what remains uncertain, and what further evidence is needed.

4.3. Illustrative Analytic Rubric: Critical Verification

Critical verification illustrates how evidence mapping can inform assessment because AI-assisted writing may introduce plausible but inaccurate claims or

citations that do not support the text. This rubric is intended for formative feedback, not as a validated scale; adapt its descriptors to the discipline, task, student level, and institutional requirements.

Table 4. Illustrative rubric for critical verification

Dimension	Level 4 — Strong	Level 3 — Competent	Level 2 — Developing	Level 1 — Limited
Identifying claims to check	Identifies and prioritizes consequential factual or interpretive claims	Identifies main claims, with minor omissions	Identifies some claims but misses important ones	Shows little recognition of claims needing verification
Checking sources and support	Uses traceable checks; evaluates credibility and context; treats unverified claims as unestablished	Uses relevant checks but incompletely evaluates context or search scope	Locates sources but pays limited attention to credibility or exact claim support	Relies on untraceable or unchecked sources, or treats AI output as verification
Responding to results	Corrects, qualifies, or removes unsupported claims and updates citations	Corrects most identified problems with minor omissions	Makes partial corrections or records problems without resolving them consistently	Leaves significant unsupported or misrepresented claims unchanged
Explaining decisions	Explains steps, evidence, decision, and remaining uncertainty	Explains main steps and decision but gives limited detail	Gives a general account with little justification	Provides no usable rationale or asserts verification without explanation

For each case, compare the original claim, the source or search used, the revised passage, and the student's rationale. A brief in-class explanation can add performance evidence. Select consequential examples rather than requiring records of every AI interaction. Rate a dimension only when evidence is sufficient; otherwise record NE (not enough evidence), not Level 1. Missing records do not necessarily indicate inability. Use ratings formatively, not as a validated total or for high-stakes decisions before piloting and checking scoring consistency. If sources conflict, seek clarification and refine ambiguous descriptors by discussing sample work.

4.4. Worked Example: A Claim That Cannot Be Verified

Consider a constructed example in which an AI tool suggests: "A 2024 study of 2,400 university students found that 68% trusted AI-generated news summaries more than journalist-written reports." The statistic and wording are invented for illustration, not a real finding. The student searches for the wording and key terms, cannot locate a traceable source supporting the statistic, removes it from the draft, and explains that failure to find a source does not prove the claim false; it means the claim cannot responsibly be presented as established evidence on the available information. The student also checks that the final paper no longer repeats it as fact.

Table 5. Illustrative Scoring of the Constructed Example

Dimension	Illustrative rating	Evidence and bounded interpretation
Identifying claims to check	NE	The student flags a precise statistic that could materially affect the argument; this shows recognition that the claim needs checking in this assignment.
Checking sources and support	Level 3 Competent	The student searches the wording and key terms, but the example provides limited evidence about search scope, source credibility, and contextual checks. The rating therefore reflects relevant but incompletely documented verification.
Responding to results	Level 4 Strong	The unsupported statistic is removed and not restated as verified fact; the student responds appropriately in this instance.
Explaining	Level 4	The student distinguishes "not verified" from "proven false" and

Dimension	Illustrative rating	Evidence and bounded interpretation
decisions	Strong	explains the evidential limit.

These ratings are illustrative, not based on actual student work; an incomplete search record could change the score. The example assesses the student's response to uncertainty, not merely whether a source was found, and supports a judgment about this decision—not a general claim of independent verification ability.

4.5. Principles for Corroborating Evidence

Corroboration compares relevant sources; it does not require every student to submit every possible record. Compare process records with the final text and revisions with the student's explanation. Check AI-use declarations against records required by the course; treat discrepancies as a reason to seek clarification, not as automatic evidence of misconduct. If uncertainty remains, use a focused oral or written explanation.

4.6. Limits of Evidence Interpretation

Process records do not provide direct access to students' thinking: interactions may not reveal understanding, and some thinking is never documented. Neither a complete portfolio nor a missing record should determine a judgment alone. Oral explanations may be affected by confidence or language proficiency, so provide accessible alternatives and assess the relevant understanding. Collect selected records of consequential decisions rather than every prompt or edit. Communicate requirements in advance and handle records under institutional privacy rules; the aim is better inference, not surveillance.

5. Illustrative Application to Undergraduate Academic Writing Instruction

This constructed case concerns a hypothetical student in an undergraduate journalism and communication course writing about how readers assess AI-generated news summaries on short-video platforms. It illustrates five stages of evidence mapping and does not describe an actual intervention.

5.1. Mapping Evidence Across the Writing Process

Stage 1: Topic development. The student moves from the broad topic of AI-generated news and university students to the narrower question of how readers judge the credibility of AI-generated news summaries on short-video platforms. A short decision note explains the topic's relevance and feasibility. Assessment focuses on the focus and rationale of the inquiry, not prompt sophistication.

Stage 2: Source evaluation and verification. While exploring concepts such as news credibility and AI-generated content, the student encounters the constructed statistic described in Section 4.4. Unable to find a traceable source supporting it, the

student excludes it and records the search and decision. The teacher compares the record with cited sources and the draft, assessing the accuracy and justification of the decision rather than the number of sources checked.

Stage 3: Argument construction. An initial AI-generated outline groups the paper around general benefits, risks, and future developments. The student reorganizes it around perceived accuracy, source transparency, and credibility cues, then explains how the revised structure supports the central question. The teacher compares the outlines, explanation, and final structure; elaboration alone is not treated as evidence of stronger reasoning.

Stage 4: Revision. The student accepts AI suggestions that improve readability but rejects or qualifies substantive claims that exceed the available evidence. A note explains two consequential changes. Comparing drafts, the note, and the final text helps distinguish surface editing from revisions affecting claims, evidence, or argument structure.

Stage 5: Submission and explanation. The student submits the paper, an AI-use declaration, selected process records, and a brief explanation of why the unsupported statistic was excluded and how the paper distinguishes perceived credibility from factual accuracy. The paper, records, and explanation provide product, process, and performance evidence. Together they support a more informed but still bounded judgment; they do not establish independent authorship or overall competence.

5.2. Evidence Corroboration in Practice

Table 6. Illustrative evidence mapping and corroboration

Outcome	Evidence	Corroboration	Permissible inference
Problem formulation	Initial and revised question; decision note	Compare the final scope with the student's rationale	The student can justify a focused writing problem in this assignment.
Critical verification	Verification log; source list; draft passage	Check source traceability and whether unsupported wording was removed or qualified	The student demonstrates a documented verification decision in this instance.
Argumentative reasoning	Outline versions; revision note; final structure	Examine alignment among the question, analytical dimensions, and conclusions	The student can explain selected structural revisions.
Ethical responsibility	AI-use declaration; selected records; final paper	Compare the declaration with required records and the student's explanation	The student demonstrates transparency about documented AI use within the available evidence.

The matrix distinguishes evidence from inference: one successful check does not establish lasting competence, nor does disclosure guarantee completeness. Link evidence to outcomes, compare sources, clarify discrepancies, and keep judgments proportionate.

6. Discussion

The framework is a context-specific operationalization of existing assessment concepts, not a new general theory. Constructive alignment informs goal–assessment links; Messick's validity perspective focuses attention on inference;

portfolio scholarship informs writing documentation; DRIVE addresses AI interactions; 4Ps distinguishes evidence types; and the four-pillar framework already combines process records, oral defense, authentic tasks, and transparent AI policies. The paper builds on these contributions.

Its contribution is a course-level map linking four adaptable writing outcomes to selected artifacts and activities through 4Ps, showing how complementary evidence supports bounded judgments. A claim is not considered verified based on a polished paper or AI log alone; instructors compare the claim, verification record, revised text, and student explanation.

DRIVE already links interaction patterns to learning objectives and reports an association between interaction quality and essay scores in STEM writing-intensive courses.[4] This framework does not claim to introduce outcome alignment; it extends the evidence map to source evaluation, outlining, revision, final writing, and student explanations. Likewise, 4Ps supplies the evidence categories rather than a new taxonomy. Because the four-pillar framework already offers operational tools, the rubric here is an illustrative design, not a novel or validated instrument.

The four outcomes are adaptable design choices, not a validated taxonomy, and should reflect course goals, genre, discipline, student level, and AI policy. As the source selection was not exhaustive and no formal review protocol, independent coding, or empirical validation was conducted, future research should examine feasibility, workload, inter-rater agreement, student perceptions of fairness and transparency, alignment with demonstrated writing capabilities, and applicability across disciplines and languages. Evidence requirements should be communicated in advance, proportionate, privacy-conscious, and learning-oriented. Conflicting evidence should prompt clarification, not automatic allegations of misconduct.

Acknowledgements

This work was supported by the 2026 Undergraduate Teaching Research and Reform Project of the University of Shanghai for Science and Technology (Grant No. JGXM202636).

References

[1] Xia, Q., Weng, X., Ouyang, F., Lin, T.J. and Chiu, T.K.F. (2024) A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, Article 40.

[2] Meng, Q., Xie, Y.-N. and Lu, D. (2026) A systematic review of AI-assisted academic writing: Tools, measures, and effects. *Journal of English for Academic Purposes*, 82, Article 101726.

[3] Lam, R. (2017) Taking stock of portfolio assessment scholarship: From research to practice. *Assessing Writing*, 31, 84–97.

[4] Oliveira, M., Zednik, C., Bombaerts, G., Sadowski, B. and Conijn, R. (2025) Assessing students' DRIVE: A framework to evaluate learning through interactions with generative AI. *Computers and Education: Artificial Intelligence*, 9, Article 100497.

[5] Fawns, T., Boud, D. and Dawson, P. (2026) Identifying what our students have learned: A framework for practical assessment validation. *Assessment & Evaluation in Higher Education*. Published online 1 February 2026.

[6] Nikolić, D. and Basta Nikolić, M. (2026) Designing AI-resilient assessment in higher education: A four-pillar conceptual framework. *Frontiers in Artificial Intelligence*, 9, Article 1841682.

[7] Biggs, J. and Tang, C. (2011) Teaching for Quality Learning at University: What the Student Does. 4th ed. Maidenhead: Open University Press.

[8] Messick, S. (1989) Validity. In: Linn, R.L. (ed.) Educational Measurement. 3rd ed. New York: Macmillan, pp. 13–103.

[9] Booth, W.C., Colomb, G.G., Williams, J.M., Bizup, J. and FitzGerald, W.T. (2016) The Craft of Research. 4th ed. Chicago, IL: University of Chicago Press.

[10] Association of College and Research Libraries (ACRL) (2026) Framework for Information Literacy for Higher Education. Revised September 2026. Available at: https://www.ala.org/acrl/standards/ilframework (Accessed: 23 October 2026).