Study of Unique Gaps and New Directions in English

(SOUGAND in English)

www.sougandinenglish.ir
ISSN: 3150-4515

Study of Unique Gaps and New Directions in English (SOUGAND in English)

Volume 1, Issue 1, 2026

REVIEW ARTICLE OPEN ACCESS

Representing Bengal: Culture, Community and Social Change in the Cinema of Tarun Majumdar

Making Authorship Visible: Assessing Generative AI in Multimodal Composing for Specific Purposes

Muhammad Shaheer Habib

School of Computing, Mathematics & Physics, University of Portsmouth, United Kingdom.
Muhammad.Habib@myport.ac.uk
ORCID: 0009-0001-6988-8138

ARTICLE INFORMATION

Received	Accepted	First Published
24 October 2026	30 October 2026	15 December 2026

DOI: https://doi.org/10.22034/sugnde.2026.256447
© 2026 The Muhammad Shaheer Habib. Published by Sougand Akbarian Publications.

Abstract

Generative AI (GenAI) is capable of generating media content like voiceovers, scripts, images, and captions, which makes it impossible to determine learners' competencies and contributions simply based on the existing multimodal products. This issue is highly relevant in the field of English and Specific Purposes Languages (ESP/LSP), as students design professional genres of communication intended for audiences that expect a high level of accuracy. This integrative-critical review explores the issue of GenAI-assisted Digital Multimodal Composing (RQ1), the challenges that come with assessing such products (RQ2), and potential principles of fair assessment of GenAI-created content (RQ3). The procedure included two rounds of online searching (39 queries in total), as well as backward citation tracing. Nineteen sources were referenced in this review, but out of 45 works read in full, only four were printed and read. The literature suggests that task design influences results, although the evidence for direct ESP/LSP outcomes still remains scarce. neither AI-use claims nor text-detection scores were capable of demonstrating any independent contribution. Recent frameworks are integrating process, product, and reflection, but there is still scarce evidence that shows whether AI use declarations are checked against something other than project results. As a solution, the author proposes the development of a provenance-based framework comparing the declaration against task-specific evidence, emphasizing the fact that corroboration cannot be considered a proof of authorship. The proposed framework has not yet been validated and does not claim any learning outcomes.

Keywords: AI declaration, assessment, authorship, digital multimodal composing, English for Specific Purposes, generative AI, provenance.

Study of Unique Gaps and New Directions in English
(SOUGAND in English)
www.sougandinenglish.ir
ISSN: 3150-4515

1. Introduction

English for Specific Purposes (ESP/LSP) equips students for workplace and academic communication. Professionals increasingly use multimodal genres, while digital multimodal composing combines writing, images, and sound for specific audiences and purposes (Jiang & Hafner, 2025; Kessler, 2024). However, ESP research remains limited (Christoforou, 2026).

Multimodal composing extends further than adding visuals to the text. Every mode brings certain tools available for generating meanings while the composer has to determine the way of modes interaction that would work for the audience (Kress, 2010). This position has a correspondence with the multiliteracies theory which presents learners as designers who utilize and transform the resources they have for meaning creation (New London Group, 1996; Cope & Kalantzis, 2009).

GenAI makes the assessment more complicated because it is capable of generating scripts, images, narration, and feedback. The aforementioned scoping review of GenAI-based DMC indicated a number of roles attributed to these instruments and revealed that the GenAI functioned as a co-composer across all twelve studies in the scope (Ci & Jiang, 2025). Thus, the refined product might provide less valuable information about the language competence, the knowledge in the discipline, and the ability to apply rhetoric in practice than was initially planned. Corbin et al. (2025) provide a recommendation as to the use of assessment tasks that would ensure that the successful accomplishment of the assignment indicates that the ability of interest has been proved.

The matter at hand is also one of accountability. A professional writer is required to take responsibility for information sources used in their work and choices of design. Theories on agency in applied linguistics indicate that both people and non-human tools can assist in the process of information making (Pennycook, 2018; Ci & Jiang, 2025).

Thus, testing in ESP/LSP should take into account the actions taken by the student as well as how he assesses and justifies these actions, not only the results of his work.

Two possible approaches to the issue include providing the range within which usage of AI is allowed, and requiring students to explain the way they used AI. The AI Assessment Scale offers levels of allowed usage (Perkins et al., 2024, 2025) while specialized forms provide students with the opportunity to elaborate on their AI usage (Micallef & Petrovska, 2026). Both options rely on rules or self-reports which may lack certain details or be misinterpreted. This leads to the question whether self-reports may be compared to the evidences obtained during the task.

A number of recent studies have begun to address DMC with the aid of GenAI technology. For example, Zhang and Li (2026) introduced a novel genre-based assessment framework involving methods of self- and peer-assessment along with human-AI composition. Further, Guo et al. (2026) compared the impact of GenAI and teachers' comments on academic posters. Shin et al. (2026) addressed the issue of assessment of outcomes, processes, and reflections in the context of ethical authorship. Related research was also conducted regarding the use of GenAI in multimodal composing by prospective TESOL teachers (Moorhouse et al, 2026). Despite the strong importance of these studies, their role in verifying the specific AI declaration needs more a comprehensive exploration.

The article offers a unique and highly thoughtful theoretical contribution. It views the AI declaration by the learner as a statement that can be verified based on task-related evidence and determines what verification of this kind can and

© 2026 Sougand Akbarian Publications | SOUGAND in English

cannot do. Additionally, this article proposes a novel provenance-based framework with examples from ESP/LSP assignments.

1.1. Key Terms

Digital multimodal composition refers to the production of digital texts that utilize language, visuals, sounds, videos, and layouts for an intended audience and goal (Jiang and Hafner, 2025; Nguyen and Nguyen, 2025).

A substantive contribution of artificial intelligence may come in forms of language, content, design, visuals, audio or layout created or altered by Generative AI and included into or influencing the object.Authorship in this instance refers to the responsibility: the learner must be able to justify the claims made by the artifact, its sources, and design choices.

Provenance refers to the established background of materials where they come from and how they were used. The process evidence is any records of the composing process, such as plans, drafts completed on dates, history of the prompts, notes for revision explanations, and others. These records, however, cannot be treated as claims because they are not complete and do not reveal the decision-maker behind every decision that has been made.

Corroboration is when statements are compared to some proof that does not come from the statements made. Though corroboration may support, clarify, or doubt a claim, it cannot be used to prove authorship, detect all issues of undeclared use, and show that no AI has been used.

The paper mentions five types of evidence in total as finished product quality, composing decisions, disciplinary knowledge, statement of the learner, and observable records of process. The summary of all of them can be found in Table 1.

Table 1

Five Kinds of Evidence and What Each Can Support

Type	Question	Typical evidence	Cannot establish
Product quality	Does the artefact meet genre, audience, and purpose criteria?	Finished artefact; teacher, expert, or peer review	Who created it or what the learner understands
Composing decisions	What choices did the learner make, and why?	Plan; annotated drafts; rationale	That choices were made without AI
Disciplinary understanding	Can the learner explain and defend the content?	Error audit; explanation; source or expert check	That understanding produced the artefact
AI-use declaration	What does the learner report AI contributed?	Mode-specific declaration	That the account is complete or accurate
Process records	What records exist for parts of the process?	Dated drafts; relevant logs; checkpoints	The entire process or who made every decision

Note. Every kind addresses a unique inquiry, and competence in one does not imply proficiency in another.

© 2026 Sougand Akbarian Publications | SOUGAND in English

2. Literature Review

2.1. Multimodality, Design and Transduction

Social semiotics considers language one of many forms of communication (Kress, 2010). In a similar vein, multiliteracies theory considers learners as designers who use the available resources to create meaning (Cope & Kalantzis, 2009). The concept of transduction, or transfer of meanings from one mode to another, illustrates how a text becomes an image, voiceover, or video (Bezemer & Kress, 2008). Although these theories help explain multi-modal meaning-making, they do not reveal the extent to which learners can be credited with that meaning-making.

2.2. Genre and Professional Communication

Genre study connects linguistic selections to disciplinary objectives and audience (Swales, 1990; Hyland, 2022). It provides a foundation for gauging if a document meets its professional context, however the specifications of the written language may ignore the effect of visual image, sound, timing, layout, and interaction. This explains why the available DMC assessment models give a good starting point for implementing genre principles to a multimodal work (Jiang et al., 2022; Jiang & Lai, 2025).

2.3. Agency, Literacy and Voice

The emergence of GenAI has brought about the debate pertaining to learner authority, voice, and critical judgement. There have been studies done related to the use of GenAI in writing which discuss various conflicts involved in creativity and standardisation including also how the learners assess the content filed by GenAI (Jeon et al., 2025; Tan et al., 2025). Other authors discuss the issues of copyright, ethics, and literacy in GenAI usage (Kostopolus, 2025; Wang & Wang, 2025). As the majority of research refers to the use of GenAI in terms of writing, application of the aforementioned studies in ESP/LSP area should be performed in a careful manner.

2.4. Assessment Validity, Permission and Provenance

According to Corbin et al. (2025), there are discursive changes that consist of instruction-based transitions which are distinct from structural changes embedded in the assessment task. Permission scale and declaration system clarify expectations but do not necessarily define how the learners have adhered to them. The modified AI Assessment Scale proposes the permitted usage to serve as a basis in dialogues and assessment (Perkins et al., 2025). Combining multiliteracies and DMC competence research can provide some criteria for assessment, although it was not aimed at verification of the GenAI involvement (Lim et al., 2024; Zhang & Yu, 2025).

In the present paper, the framework is based on a simple premise: exposing AI usage, assessing the output and assessing the process of composing the work. The task-specific declaration is more informative when a teacher is able to verify specific information against appropriate evidence from the task. The evidences from computer education proves that declarations and workflow should be compared but not assumed (Micallef & Petrovska, 2026). This review does not discuss how often such mismatches are present in ESP/LSP.

2.5. What Existing Work Establishes

In the literature analyzed, three main trends can be observed. Firstly, results depend on the nature of the task and teacher assistance (Ci & Jiang, 2025; Nguyen & Nguyen, 2025). Secondly, human evaluation is necessary to assess the purpose, substance, and context of multimodal products; although GenAI assessments may give helpful feedback

© 2026 Sougand Akbarian Publications | SOUGAND in English

concerning posters, it does not satisfy every criterion (Guo et al., 2026). Lastly, claims about critical thinking and authorship involve students' responses or reflections, and detection studies reveal considerable limitations and threats to multilingual authors (Liang et al., 2023; Weber-Wulff et al., 2023).

The data vary depending on the task and context, which means that broad statements regarding GenAI's advantages or disadvantages should be approached with caution.

In Table 2, the closest evaluation methods discovered in the accessible literature are compared. The identified limitations refer only to the examined literature rather than serve as evidence for the existence of no other relevant publications.

Table 2

Closest Existing Work and What It Leaves Open

Approach	What it contributes	Open question in accessible evidence	Use in this article
Process–genre model (Jiang & Lai, 2025)	Links GenAI-assisted process with product quality	Does not clearly test declarations against records	Process–product relationship
Genre-based GenAI DMC model (Zhang & Li, 2026)	Refines criteria through self/peer assessment; adds human–AI process	ESP/LSP-specific corroboration is not established in the abstract	Genre criteria and process layer
Poster feedback study (Guo et al., 2026)	Compares GPT-4o and teacher feedback on 88 posters	Static products; declaration checking not reported	Human–AI division of assessment
Ethical authorship assessment (Shin et al., 2026)	Connects product, process, and reflection	Requires closer full-text comparison with provenance checking	Closest conceptual precedent
Process assessment (Cheung, 2023; Hafner & Ho, 2020)	Provides process/product criteria for DMC	Predates widespread GenAI use	Process evidence
AI permission scale (Perkins et al., 2024, 2025)	Clarifies permitted levels of AI use	Not specific to multimodal provenance	Permission layer

© 2026 Sougand Akbarian Publications | SOUGAND in English

Approach	What it contributes	Open question in accessible evidence	Use in this article
Task-specific declarations (Micallef & Petrovska, 2026)	Offers activity categories and examples	Self-reports need workflow verification.	Declaration template
Structural assessment (Corbin et al., 2025)	Distinguishes instructions from task-embedded changes	Not specific to ESP/LSP DMC	Built-in capture points
Critical digital literacies (Hafner & Ho, 2026)	Frames agency, creativity, and criticality	Not primarily an assessment model	Agency dimension
AI literacy cycle (Nguyen & Nguyen, 2025)	Organises interpreting, generating, evaluating, and revising	Not validated as a provenance assessment	Task sequence

Note. Note: Entries include either abstracts or text that paper is open and accessible.

2.6. The Gap, Narrowed

The analysis failed to find a methodology that combines the evaluation of mode-specific and stage-specific declarations, assignment-produced evidence, or the characteristics of genre in ESP/LSP artifacts. This is a temporary gap rather than a statement of uniqueness since the frameworks presented by Shin et al. (2026) might coincide with some parts of the proposal. The important contribution of the framework might be the ability to identify the claims as consistent, partly consistent, inconsistent, or untestable, thus eliminating the inability to treat missing or contradicting evidence as evidence of misconduct.

2.7. What Is New and What Extends Existing Work

The contribution aims to augment existing work, and is not an attempt at replacing the preceding one. The article categorically separates product quality, learner choices, discipline understanding, declarations of deployments of AI, and records of processes. It elaborates the limits of verification and introduces rules for decisions and relevant tasks applicable in ESP/LSP. However, the framework remains a proposal for evaluation rather than a validated model.

The contribution might turn out to be more limited in case the previous framework contains specifications on how reflections are verified and supported by processing evidence. Therefore, one should analyze Shin et al. (2026) in detail to be sure of higher levels of novelty.

This review answers three research questions:

© 2026 Sougand Akbarian Publications | SOUGAND in English

RQ1. What does recent research reveal about GenAI-supported DMC in ESP/LSP and similar contexts of language education?

RQ2. What problems arise with assessment of AI-supported multimodal artifacts in the area of professional and disciplinary communication?

RQ3. Which knowledge-based principles can help conduct fair ESP/LSP assessment considering the involvement of AI, learner's understanding, accessibility, and discipline requirements?

3. Methodology

3.1. Design

The current research employs a critical integrative review process to integrate and summarize works conducted in the fields of language education and writing studies as well as assessment (Whittemore & Knafl, 2005; Torraco, 2005, 2016). It was a single researcher who did the review. This review cannot be referred to as systematic or scoping because there were no protocols utilized or databases searched and screening counts were not retained. So, this research is conducted in a somewhat exploratory way and the research gap claim is tentative.

Table 3

Review Protocol as Applied

Element	What was done
Design	Critical integrative review; one researcher; not systematic or scoping
Search	39 general web queries in two phases (16 and 23 October 2026), plus backward citation checks; no direct database searches
Eligibility and screening	Relevant to GenAI, DMC, ESP/LSP, assessment, disclosure, integrity, equity, or teacher development; informal single-researcher screening; counts not recorded
Reading depth	Of the 45 cited works, 4 were read in full, 13 informed the background, and the rest were reviewed through abstracts, excerpts, or records.
Appraisal and extraction	Evidence matrix covering context, design, findings, limitations, and relevance; criteria adapted by source type
Synthesis	Thematic analysis organised by research question; framework compared with nearest models

Note. Publisher pages on Wiley, Taylor & Francis, and ScienceDirect were inaccessible during Phase 2.

© 2026 Sougand Akbarian Publications | SOUGAND in English

3.2. Search and Selection

On October 16, 2026, the initial phase of the search was conducted and comprised of 24 queries spread over four rounds which revolved around topics such as GenAI-assisted DMC, ESP/LSP, AI approval and declaration systems, detector bias. Additionally, it included multimodality, and review techniques. Two references were cited and backward citation analysis was performed. Through the search database, entire documents, sources and aggregators were checked for their validity through titles and abstracts.

On October 23, 2026, a second round of searches was performed during which 15 queries were run about linguistic assessment, ethical authorship, and bibliographic verification conducted from 2025 to 2026. But databases like ERIC, Crossref, OpenAlex, Google Scholar, Scopus, and Web of Science were not searched in this phase. Due to access restrictions of some publisher sites, some sources were analyzed from their abstracts, excerpts or cataloged records. Before completing this phase, Moorhouse et al., 2026 was included in the last round of references searched for GenAI-assisted DMC in TESOL teacher education.

3.3. Appraisal, Synthesis and Reading Depth

The academic reference framework documented the context, designs, findings, limitations, and relevance of each source. Appraisal included an assessment of source type, transparency, sample, and match to the research questions. There is no single score awarded for quality.

The references list includes 45 works. Four sources were read in full including Christoforou (2026), and the articles from Ci and Jiang (2025), Micallef and Petrovska (2026), and Nguyen and Nguyen (2025). Thirteen conceptual or methodological sources were consulted as background rather than primary. Other sources were examined only in varying levels of details using abstracts, excerpts, or bibliographic data. Five studies are discussed only as cited by Micallef and Petrovska (2026), so they are treated as a secondary source rather than an independent finding.

The limitations of this review include lack of any information regarding the database coverage, lack of reading any of the sources completely, and having a single investigator who did the material screening without recording how many items were screened. The findings can suffer from publication bias and language bias, and the findings relating to general L2 context may not be applied to the ESP/LSP.

4. Results

The results are structured according to the research questions. After that, the proposed approach for addressing RQ3 is presented.

4.1. The Evidence Base

The research on GenAI-supported DMC is new and mainly focused on university students in language classes. In the scoping review conducted by Ci and Jiang (2025), twelve empirical research studies were counted. The authors Nguyen and Nguyen (2025) have discovered seven AI-driven frameworks out of the 48 DMC frameworks. The studies encompass such areas as storytelling, videos, posters, children's books, composing by preservice teachers and cognitive activities. The research on ESP is quite scarce, with a minor case study on English for Chemical Engineering by Christoforou (2026) and research conducted by Hafner and Ho (2020, 2026) on digital multi-modal composing in

© 2026 Sougand Akbarian Publications | SOUGAND in English

English for scientific purposes. The preservice TESOL teacher education study by Moorhouse et al. (2026) provides some additional information but does not answer the question about the validation of task-dependent statements.

Results from general L2 contexts may not apply similarly in high-stakes professional contexts.

4.2. The Role of GenAI and Task Design

Jiang and Lai (2025) analyzed 14 GenAI-assisted groups and 10 non-GenAI groups across four English classrooms. The groups using GenAI performed better on target development, planning/navigating, and overall results, while some other aspects were similar. The abstract indicates that it is possible that the groups may have opted for GenAI, hence making it difficult to attribute results solely to the tool. It has also been shown that results are determined partly by the design of the task and scaffolding provided (Ci & Jiang, 2025; Nguyen & Nguyen, 2025).

4.3. Modes and Limits of AI Assessments

There are different ways in which GenAI assessment and feedback may differ. In previous DMC research, limitations of assessing coherence, narrativity, and style were highlighted, leaving teachers and students in charge of multimodal assessments (Ci & Jiang, 2025). Guo et al. (2026) compared assessments made by GPT-4o and a teacher of posters created by 88 Chinese university EFL students. According to the findings of the study, AI assessment agreed with some criteria, but the role of human judgment remained important.

The case study of Chemical Engineering showcases the importance of expert judgement; the case describes a situation where a student rejected industrial images produced by artificial intelligence simply because they were not accurate scientifically, and presented her own reasoning (Christoforou, 2026). The implications of this for evaluation is that provenance should be described in a particular mode and that visual appeal should be distinguished from correctness.

Critical engagement is not something one should assume based on a self-report; there should be clear evidence in terms of task requirements.

4.4. Genre and Professional Communication

Students can refer to popular digital genres like how-to videos in order to demonstrate professionalism but such conventions may not reflect workplace genres accurately. Studies on science documentary making show the interplay of modes, steps, outcomes, voice, and authenticity (Hafner & Ho, 2026). The sources analyzed show that students usually reported disciplinary accuracy, but did not check it with the help of experts.

4.5. Agency, Criticality and AI Literacy

Research regarding AI literacy illustrates the importance of enhancing assessments instead of presupposing it. For instance, results from Nguyen and Nguyen (2025) do not show significant benefits of AI-based approaches to critical thinking. In the ESP case, students noticed some irrelevant outputs, but a limited sample of purposeful selection and teacher-researcher role restricts the inferences made (Christoforou, 2026). Despite the relevance of the critical AI literacy models and intellectual property discussion, most of the studies are related to writing in general rather than multimodal professional products (Kostopolus, 2025; Wang & Wang, 2025).

© 2026 Sougand Akbarian Publications | SOUGAND in English

4.6. Models of Assessment

DMC assessments based on process have been around even before the advent of GenAI. Hafner and Ho (2020) explain that formative and summative assessment includes stages of planning, design, sharing, and reflecting stage; Cheung (2023) elaborates rubrics regarding process and product. Some studies have recently combined ideas of process and product (Jiang & Lai, 2025), while others improved genre assessment criteria using self and peer assessment (Zhang & Li, 2026), and introduced a greater emphasis on product, process, and reflection in relation to ethical authorship (Shin et al., 2026). These studies are examples of early predecessors.

4.7. Disclosure, Integrity and Fairness

Disclosure does not ensure transparency. Evidence shows that disclosures may not be comprehensive, misunderstood, or distorted by the fear of consequences (Micallef & Petrovska, 2026). Other sources explain that detection mechanisms are inefficient and that multilingual authors are unfairly accused (Liang et al., 2023; Weber-Wulff et al., 2023).

According to Micallef and Petrovska (2026), studies reveal that declarations are not exhaustive, AI usage is minimally described, and there is uncertainty about what applies to GenAI. The figures are mentioned as secondhand evidence because such studies have not been consulted independently. Evidence proves that text detection is useless. Liang et al. (2023) reported the risk of significant false positive results for the works written by non-native speakers of English, while Weber-Wulff et al. (2023) mentioned a low reliability of detection methods. The literature did not provide evidence of detection processes for visual or audio-based works. Fair evaluation should also consider issues of accessibility, confidentiality, and disability and whether the imposed record keeping leads to disproportionate inconvenience (Jeon et al., 2025; Ci & Jiang, 2025)

4.8. Patterns Across the Themes

Three patterns cut across the review. First, studies describe GenAI's role more often than they measure learners' capability without AI support. Second, claims about agency and criticality frequently rely on interviews, reflections, or researcher interpretation. Third, existing assessment approaches include product, process, and reflection, but the accessible evidence offers little direct testing of whether these sources distinguish learner contribution from AI contribution. Most evidence is qualitative, and findings from general L2 contexts may not transfer directly to professional ESP/LSP tasks.

4.9. Suggested Framework

The suggested framework addresses research question 3 through aligning the genre-based product criteria with the evidence on decisions, understanding, declarations, and task records.

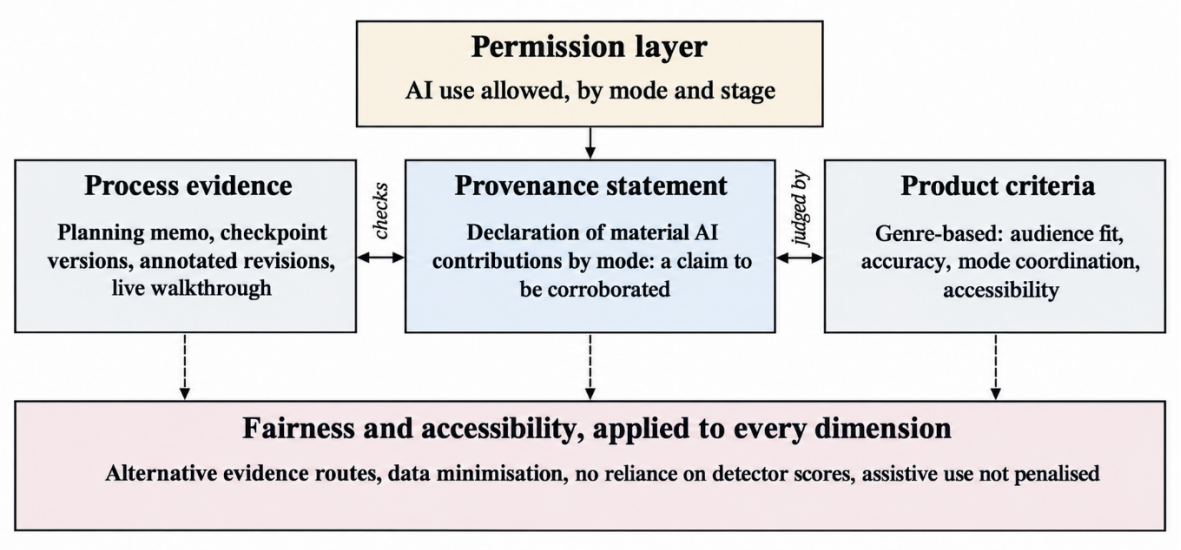
4.10. Elements and Principles

The framework relies on four points: contribution of AI can vary by mode, human judgement is still critical for purpose and content (Guo et al., 2026), declarations lack perfection, and task design allows for producing evidence beyond that described by the instructions (Corbin et al., 2025). Figure 1 provides the framework in question, while Table 3 lists six assessment aspects.

© 2026 Sougand Akbarian Publications | SOUGAND in English

Figure 1

Provenance-Based Assessment Framework



Note. The framework links AI-use rules, assessment components, and fairness, but remains unvalidated.

Table 4

Proposed Assessment Alignment Matrix and Rubric Structure (Unvalidated)

Dimension	Assessment question	Primary evidence	Learning outcome	Risk and safeguard
Purpose, audience, and genre	Does the artefact suit its professional purpose?	Brief, artefact, genre criteria, audience response	Communicate to a defined audience	Tool conventions may replace field genres; consult subject specialists
Disciplinary accuracy and language	Are content, terminology, and register accurate?	Artefact; source or expert check	Use accurate disciplinary content	Fluent errors may pass unnoticed; verify against sources
Multimodal coherence and accessibility	Do the modes work together and remain accessible?	Artefact; captions, alt text, readability, contrast	Coordinate modes inclusively	Polish may mask poor coordination; judge against purpose
Decisions and revision	How did the learner evaluate and revise AI output?	Plan; checkpoints; annotated revisions; error audit	Evaluate and revise critically	Records may be AI-assisted; accept varied traces and avoid over-weighting one

© 2026 Sougand Akbarian Publications | SOUGAND in English

ISSN: 3150-4515

Dimension	Assessment question	Primary evidence	Learning outcome	Risk and safeguard
Understanding of content and design	Can the learner explain and defend key choices?	Walkthrough, questions, written or interpreted alternative	Defend decisions to an audience	Anxiety or disability may affect oral response; offer equivalent routes
Provenance consistency	Do declaration and available records align claim by claim?	Mode-specific declaration compared with relevant records/explanation	Account for AI contributions	Avoid surveillance and penalising permitted use; mark unsupported claims not testable

Note. The principles of fairness and accessibility apply in collecting evidence but are not included in the assessment. All pieces of evidence are scored for just one fundamental principle.

To avoid double counting, every piece of evidence is allocated to one principle only: the principles guide actions; an explanation gives insight; and assertions are checked against recorded information regarding their origin. Evidence may generate questions related to other principles but should not be rewarded additionally. AI usage is not penalised in formulating a framework, as only the terms used and learner's achievements are assessed without Alice's contribution.

The principles discussed further are proposals developed from research rather than interventions that were tested.

- Focus tasks around the genre and the audience in mind; and derive the criteria from the purposes of the discipline rather than from the limitations of the tool.
- Identify the allowed modes of AI usage because there is a huge difference between producing narratives and offering some layout ideas.Build proportionate evidence points into the task, such as a plan, a draft checkpoint, and an explanation.
- Compare each material declaration with available task evidence; label claims that cannot be checked as not testable.
- Match the amount of evidence collected to the stakes of the task; formative work needs lighter documentation.
- Treat discrepancies as a prompt for discussion, since they may reflect forgetting, tool ambiguity, or misunderstanding (Micallef & Petrovska, 2026).

4.11. What Corroboration Can and Cannot Establish

Corroboration indicates whether there is evidence that supports, partially supports, or conflicts with the material claim. The result is designated as in agreement, partly in agreement, not in agreement, or not able to be tested. Not able to test only describes the evidence and not the integrity of the student.

The limitations are huge. The process traces may be incomplete or produced by machine learning; different platforms may have different information; the live presentation may be scripted. The level of certainty is improved by consistency but does not prove authorship. Mismatch is a reason for discussion and further investigation and does

© 2026 Sougand Akbarian Publications | SOUGAND in English

not mean that there is automatically academic dishonesty. Without task structures producing the evidence, the claims remain purely self-reports.

4.12. An Operational Sequence for Teachers

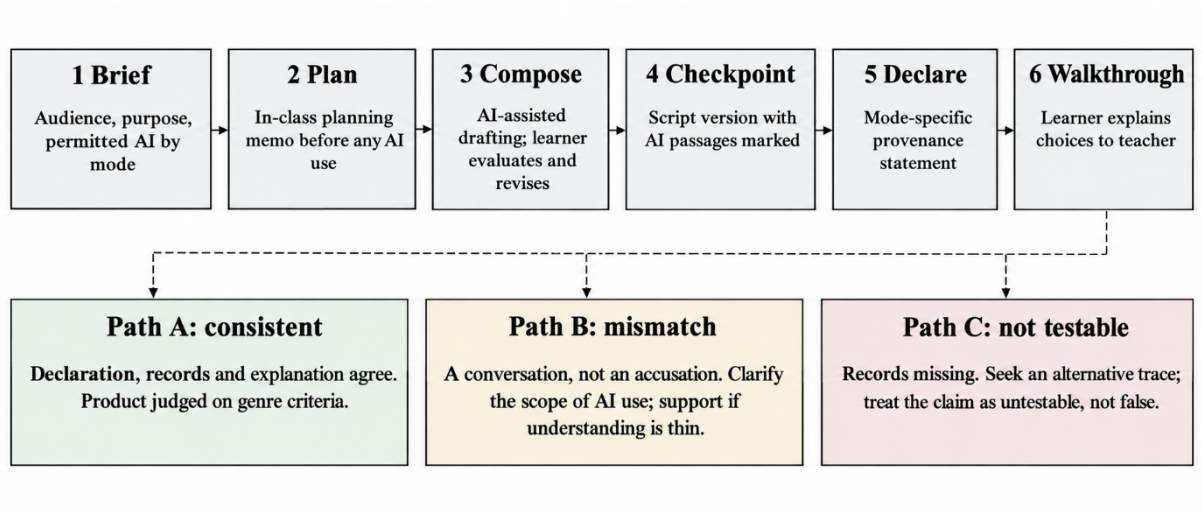
The proposed sequence has six steps. First, define permitted AI use by mode and stage, distinguishing access-related assistance from content generation. Second, ask learners for a mode-specific declaration covering activity type, extent, examples, and how outputs were used. Third, select two or three proportionate capture points and offer accessible alternatives. Fourth, assess the artefact against genre criteria, assess decisions and understanding from relevant evidence, and compare each declaration claim with available records. Fifth, discuss missing or conflicting evidence with the learner and record the reasoning. Sixth, moderate a sample of judgements and revise the task and permissions. Co-constructed criteria may support transparency (Shin et al., 2026).

4.13. A Worked Example

As an illustrative proposal, a learner in an English for Chemical Engineering course creates a three-minute video explaining a batch process to novice peers. AI use is permitted for a script outline, caption drafts, and background images; process steps and terminology must be checked against course sources. Three capture points are planned: an in-class memo before AI use, a script checkpoint with AI-origin passages marked, and a short walkthrough in which the learner explains key choices.

Figure 2

Worked Example: Task Sequence with Capture Points and Three Possible Outcomes



Note. Illustrative proposal, not a reported case.

The learner states limited AI usage in planning and moderate assistance for revised captions while relying heavily on AI for generating background images. Two images are substituted since they inaccurately depict used equipment. The draft agrees with the declaration, and the artefact is evaluated based on its precision, appropriateness for the audience, multimodal synergy, and captioning.

© 2026 Sougand Akbarian Publications | SOUGAND in English

The next case sees the learner stating very little text-utilizing. In this instance, unchanged paragraphs are retroactively included in the draft, causing confusion that does not necessarily imply misconduct on the part of the learner. Discussion may lessen the ambiguity and provide information on whether it was a misunderstanding regarding auto-completion or whether the learner needs assistance with the terminology.

The third case features the composing software containing no evidence of prior work. The teacher supplies other types of evidence where applicable such as date-stamped files or annotated writings, and records the declaration as unproven, rather than incorrect if no proof can be obtained.

4.14. Handling Missing or Contradictory Evidence

Table 4 helps one to take decisions about the absence or inconsistency of evidence. Pieces of evidence may lack completeness because of the limits of a platform, unequal access, disability, anxiety, or ambiguity of instructions. Thus, teachers have to clarify with learners, give alternative means of assessment, and make a judgment based solely on the piece of evidence presented.

Table 5

How to Handle Absence or Contradictory Evidence

Situation	Possible explanations	Reasonable response	Avoid
Declaration, records, and explanation agree	Account may be consistent	Record consistency; assess product on genre criteria	Treating consistency as proof of authorship
Declaration understates use	Forgetting, unclear categories, translation/autocomplete, or omission	Discuss records, clarify terms, invite correction; assess understanding separately	Inferring misconduct from mismatch alone
Extensive use declared but understanding is weak	Learning need, anxiety, language barriers, or lack of preparation	Offer accessible alternative; give feedback; score only demonstrated understanding	Penalizing permitted AI use
Records are missing	Platform limits, lost files, unequal access	Accept an alternative trace; mark claim not testable if necessary	Treating missing records as dishonesty
Records conflict	Different tools/accounts or incomplete logs	Ask for explanation; use more direct records and a second marker for high stakes	Selecting only evidence that fits suspicion

© 2026 Sougand Akbarian Publications | SOUGAND in English

Situation	Possible explanations	Reasonable response	Avoid
Learner declines process-data sharing	Privacy or accessibility concern	Offer a less intrusive route and explain data handling	Requiring full logs or screen recording by default
Detector score is cited	Detector reliability and fairness are limited	Set the score aside; return to task evidence	Using a detector to decide a case
Independent evidence suggests serious misrepresentation	Possible integrity concern	Follow institutional procedure; document evidence and hear the learner	Informal sanctions outside procedure

Note. The guide supports judgement and does not replace institutional integrity procedures.

4.15. Applications and Assessment

The following applications illustrate possible uses of the framework and the trade-offs of different evidence sources.

4.16. Illustrative Tasks

Table 5 presents three proposed tasks for discussion; none has been tested in a classroom study.

Table 6

Illustrative Multimodal Tasks (Proposals, Not Reported Studies)

Element	A. Procedural video	B. Visual research summary	C. Narrated presentation with questions
Format	Time-based	Static	Time-based
Audience and purpose	Explain a process to novice peers	Report findings to non-specialists	Defend a proposal to an expert panel
Modes	Speech, text, images, video	Text, data visualization, layout	Slides, speech, diagrams
Permitted AI	Script outline, caption drafts, background images	Plain-language edits, layout ideas, alt-text drafts	Language feedback and translation checks
Learner decisions	Check steps and images for accuracy	Check claims against sources; decide what to omit	Select evidence and respond to questions

© 2026 Sougand Akbarian Publications | SOUGAND in English

Element	A. Procedural video	B. Visual research summary	C. Narrated presentation with questions
Process evidence	Plan, annotated script, walkthrough	Source-to-claim table, version history, explanation	Live questions or equivalent accessible check
Assessment criteria	Accuracy, audience fit, coordination, captions	Fidelity, clarity, accessibility, data display	Argument, evidence, delivery, responsiveness
Main risk	Plausible but incorrect generated material	Simplification distorts findings	Rehearsed answers or anxiety

Note. None of these tasks has been tested.

4.17. Evaluating AI Output

Learners require practice in assessing and rectifying the outputs of AI. The activities comprise spotting errors against relevant literature, investigating stereotypes and registering any inaccuracies, including errors in register and representation (Nguyen & Nguyen, 2025; Christoforou, 2026). The error audit sheet can highlight revision choices. AI preparation encourages learner's initial intentions, though preparatory notes or marked-up rewrites do not prove that the work was plagiarism-free.

4.18. Process Evidence and Its Trade-offs

The records of various processes illustrate different aspects of writing process (Table 6). Drafts and version histories show the changes made but depend on the technologies being used. Comments help the revision process but can still be completed due to AI. Portfolios show the evolution of the knowledge but increase the amount of work to be done; comments from peers help in making decisions about evaluation but differ in quality.

Table 7.

Sources of Process Evidence: What Each Can and Cannot Show

Source	Can show	Cannot show	Main risk	Safeguard
In-class plan before AI	Initial ideas and intent	No AI use outside class	Rehearsal even in low stakes	Keep short and lightly weighted
Dated drafts/checkpoints	How artefact changed	Who made every change	Platform dependence, unequal access	Accept multiple formats
Prompt/chat export	Logged prompts and outputs	Other tools/accounts or edited exports	Privacy, incomplete logs	Share relevant excerpts only; delete after cycle

© 2026 Sougand Akbarian Publications | SOUGAND in English

Source	Can show	Cannot show	Main risk	Safeguard
Annotated revisions/error audit	Reasoning about corrections	Unaided reasoning	May be AI-assisted	Pair with another proportionate source
Walkthrough/questions	Understanding and defense of choices	Who made every decision	Anxiety, language/disability barriers	Consistent prompts and alternatives
Observed checkpoint	Work in progress	The rest of the process	Time and uneven coverage	Use few planned checkpoints
Screen recording	One observed session	Work in other sessions/devices	Surveillance and access burden	Not default; consent and alternative
AI-detector output	No reliable authorship evidence	Authorship	Bias and untested multimodal use	Exclude from authorship decisions

Note. Reliability depends on whether the task generates the record or the learner supplies it afterwards.

4.19. Declarations, Detection and Privacy

Despite the fact that declarations foster transparency, they cannot provide proofs of authorship as students can misreport the use of AI systems and rely on secondary evidence (Micallef & Petrovska, 2026). Results of detectors cannot substitute for evidence of performance and can be unfair to non-native authors (Liang et al., 2023; Weber-Wulff et al., 2023). The current state of multimodal detection still leaves much to be desired.

Excessive process gathering may amount to surveillance. Limit collected data to what is necessary, state its purpose, and get rid of all information after the assessment. Rather than mandatory screen recordings, permit sketches, audio notes, manuals, or verbal assessments. Accept assisted or authorized AI use. Organizations should provide equal access to relevant technologies and clear regulations about privacy issues.

4.20. Rubric Development and Validation

Prior to utilizing the rubric for high-stakes evaluation, it is crucial to validate it through expert content review, teacher think-alouds research, inter-rater reliability assessments, and examination of possible variations in outcomes experienced by multilingual students and students with disabilities.

5. Discussion

5.1. Answering the Research Questions

The review of literature provides temporary answers to three questions that have formed the foundation of this research.

© 2026 Sougand Akbarian Publications | SOUGAND in English

RQ1. The work concerned GenAI-assisted DMC being a collaborative effort that relies on roles played by tools and the design of a task. The research is mostly limited to language education in undergraduate settings while not much attention is paid to ESP/LSP studies (Ci & Jiang, 2025; Christoforou, 2026; Hafner & Ho, 2026; Moorhouse et al., 2026).

RQ2. The assessment of an artifact does not answer the question of how well the learner has worked. Human judgement still matters, as discipline purpose and content need to be taken into account as declarations can also miss out information while the use of detectors may not be successful due to inaccuracy in multilingual cases (Corbin et al., 2025; Guo et al., 2026; Liang et al., 2023).

RQ3. Six provisional principles have been elaborated: align tasks according to genre and audience; indicate AI use in terms of mode and stage; gather appropriate evidence; juxtapose declarations and available information; assess claims that cannot be tested objectively; discuss misunderstandings. The principles need to be discussed in terms of fairness and accessibility of the framework.

5.2. Implications for Teachers and Institutions

Educators must possess the literacy of GenAI, genre-specific standards, and skills in task design. Human judgement is necessary because AI feedback is not always reliable (Ci & Jiang, 2025; Guo et al., 2026). The process of developing professionally can help attain these capabilities (Christoforou, 2026).

Accessibility can be considered a measure of communicative quality. Captions, alternative text, readable content, and equal means for expressing design choices should be accessible for users. Provenance templates should help differentiate the uses of an assistive technology from the content generated by means of AI. Such an approach must be tested.

Educators must clarify guidelines on the use of AI, outline significant capture points, and discuss requirements with students. Designers must also refer to experts in the field during the creation of the rubrics and test them before using them for high-stakes purposes. Institutions must align their policies with task rules, avoid decisions based solely on detection, provide equal access to the tool, and safeguard process data.

6. Conclusion

6.1. Limitations

The review has its limitations. It employed a generic search tool instead of a database search, it did not take screening counts, and it only reviewed four out of the 45 references in their entirety. Some key literature including recent pieces that may be able to narrow the gap was only reviewed through abstracts, excerpts or records. Only one researcher was involved in the review. Given the limited amounts of ESP/LSP-specific research, the current framework partly relies on general L2 literature. Its practicality, construct validity, reliability, impartiality, and transferability have yet to be determined.

6.2. Suggestions for Further Research

The framework prompts the following questions that can be examined:

© 2026 Sougand Akbarian Publications | SOUGAND in English

- Is the presence of the capture points likely to enhance the consistency between claims and actual workflow observed? If not, the framework's argument is weak.
- Are it possible to predict actual performance based on understanding and provenance scores? If the answer is negative, it means these scores do not measure learning.
- Is there any variability in mismatch flags, scores, and workloads among multilingual learners and learners with disabilities?
- Is the framework proportional? The time needed to execute tasks, perceived surveillance, and behavioral effects should be assessed.
- Is the process evidence unique and is it possible to find the explanation that provides useful evidence with minimal intervention?
- Is the framework applicable in the fields of LSP/ESP and does it work in institutions of varying wealth levels?

If the findings of validation don't show great added value compared to the practices that already exist, then the contribution should focus on ESP/LSP advice and guidelines pertaining to specific genres.

6.3. Final Remarks

In ESP/LSP, learners were to be more accurate and assertive in explaining their choices. This means that any contribution by GenAI towards the making of a multimodal artefact by learners does not verify learner participation by itself. For proportionate evidence, a participant can clarify the decisions and declarations, but declarations do not substantiate authorship. The framework needs to be tested empirically to determine if it is effective and fair.

References

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195. https://doi.org/10.1177/0741088307313177

Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing, 56*, Article 100723. https://doi.org/10.1016/j.asw.2023.100723

Christoforou, M. (2026). Gen AI-assisted multimodal meaning design: Exercising a pedagogic metalanguage of transposition. *Pedagogies: An International Journal, 21*(2), 276–300. https://doi.org/10.1080/1554480X.2025.2522884

Ci, F., & Jiang, L. (2025). The integration of generative artificial intelligence into digital multimodal composing in second language classrooms: A scoping review from the perspective of tasks. *Digital Applied Linguistics, 2*, Article 102939. https://doi.org/10.29140/dal.v2.102939

Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal, 4*(3), 164–195. https://doi.org/10.1080/15544800903076044

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education, 50*(7), 1087–1097. https://doi.org/10.1080/02602938.2025.2503964

© 2026 Sougand Akbarian Publications | SOUGAND in English

Guo, K., Li, D., & Jiang, L. (2026). Evaluating generative artificial intelligence's role in enhancing feedback for digital multimodal compositions: A comparative study. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2026.2640476

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Hafner, C. A., & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing, 47*, Article 100710. https://doi.org/10.1016/j.jslw.2020.100710

Hafner, C. A., & Ho, W. Y. J. (2026). Doing digital multimodal composing with generative artificial intelligence: Towards an integrated critical digital literacies model. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688261459662

Hyland, K. (2022). English for specific purposes: What is it and where is it taking us? *ESP Today, 10*(2), 202–220. https://doi.org/10.18485/esptoday.2022.10.2.1

Jeon, J., Wei, L., Tai, K. W. H., & Lee, S. (2025). Generative AI and its dilemmas: Exploring AI from a translanguaging perspective. *Applied Linguistics, 46*(4), 709–717. https://doi.org/10.1093/applin/amaf049

Jiang, L., & Hafner, C. A. (2025). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching, 58*(4), 528–546. https://doi.org/10.1017/S0261444824000107

Jiang, L., & Lai, C. (2025). How did the generative artificial intelligence-assisted digital multimodal composing process facilitate the production of quality digital multimodal compositions: Toward a process-genre integrated model. *TESOL Quarterly, 59*(S1), S52–S85. https://doi.org/10.1002/tesq.3390

Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing, 57*, Article 100869. https://doi.org/10.1016/j.jslw.2022.100869

Kang, J., & Yi, Y. (2023). Beyond ChatGPT: Multimodal generative AI for L2 writers. *Journal of Second Language Writing, 62*, Article 101070. https://doi.org/10.1016/j.jslw.2023.101070

Kern, R. (2024). Twenty-first century technologies and language education: Charting a path forward. *The Modern Language Journal, 108*(2), 515–533. https://doi.org/10.1111/modl.12924

Kessler, M. (2024). *Digital multimodal composing: Connecting theory, research and practice in second language acquisition*. Multilingual Matters.

Kostopolus, E. (2025). Student use of generative AI as a composing process supplement: Concerns for intellectual property and academic honesty. *Computers and Composition, 75*, Article 102894. https://doi.org/10.1016/j.compcom.2024.102894

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

© 2026 Sougand Akbarian Publications | SOUGAND in English

Li, M., & Akoto, M. (2021). Review of recent research on L2 digital multimodal composing. *International Journal of Computer-Assisted Language Learning and Teaching, 11*(3), 1–16. https://doi.org/10.4018/IJCALLT.2021070101

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lim, F. V., Toh, W., & Ong, M. (2024). Assessing multiliteracies: A systematic review and a perspective from Singapore. *Asia Pacific Journal of Education, 44*(3), 597–610. https://doi.org/10.1080/02188791.2024.2390658

Lin, C.-H., Zhou, K., Li, L., & Sun, L. (2025). Integrating generative AI into digital multimodal composition: A study of multicultural second-language classrooms. *Computers and Composition, 75*, Article 102895. https://doi.org/10.1016/j.compcom.2024.102895

Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education, 211*, Article 104977. https://doi.org/10.1016/j.compedu.2023.104977

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Micallef, N., & Petrovska, O. (2026). *Structuring transparency: Developing domain-specific generative AI declaration frameworks in higher education* [Preprint]. arXiv. https://arxiv.org/abs/2606.13389

Moorhouse, B. L., Hafner, C. A., & Ho, T. Y. (2026). GenAI-assisted digital multimodal composing in TESOL teacher education. *TESOL Quarterly*. Advance online publication. https://doi.org/10.1002/tesq.70184

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Nguyen, T. T. T., & Nguyen, T. N. Q. (2025). AI literacy in language learning through digital multimodal composing: A systematic review. *VNU Journal of Foreign Studies, 41*(3), 142–156. https://doi.org/10.63023/2525-2445/jfs.ulis.5517

Pennycook, A. (2018). *Posthumanist applied linguistics*. Routledge.

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(6), Article 06. https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., & Furze, L. (2025). Reimagining the Artificial Intelligence Assessment Scale (AIAS): A refined framework for educational assessment. *Journal of University Teaching and Learning Practice, 22*(7). https://doi.org/10.53761/rrm4y757

Shin, D.-s., Yi, Y., & Cimasko, T. (2026). Assessment of digital multimodal composing: Toward ethical and responsible use of generative AI. *TESOL Quarterly*. Advance online publication. https://doi.org/10.1002/tesq.70165

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

© 2026 Sougand Akbarian Publications | SOUGAND in English

Tan, X., Xu, W., & Wang, C. (2025). Purposeful remixing with generative AI: Constructing designer voice in multimodal composing. *Computers and Composition, 75*, Article 102893. https://doi.org/10.1016/j.compcom.2024.102893

Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies, 6*(3), 36–57. https://doi.org/10.58304/ijts.20240304

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67*, Article 101187. https://doi.org/10.1016/j.jslw.2025.101187

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Zhang, E. D., & Yu, S. (2025). The development and validation of an L2 student digital multimodal composing competence scale. *Computer Assisted Language Learning, 38*(3), 486–511. https://doi.org/10.1080/09588221.2023.2201342

Zhang, M., Akoto, M., & Li, M. (2023). Digital multimodal composing in post-secondary L2 settings: A review of the empirical landscape. *Computer Assisted Language Learning, 36*(4), 694–721. https://doi.org/10.1080/09588221.2021.1942068

Zhang, Y., & Li, Z. (2026). Assessing GenAI-assisted digital multimodal composing: Reconceptualizing a genre-based framework through self-assessment and peer assessment. *Assessing Writing, 68*, Article 101017. https://doi.org/10.1016/j.asw.2026.101017

Acknowledgements

Generative AI (Claude, Anthropic) played a crucial role in conducting literature searches, verifying sources, structuring, writing, and proofreading the manuscript, even in October 2026. The author approved of the manuscript and is accountable for its content and citations.

Conflict of Interest

There is no conflict of interest.

© 2026 Sougand Akbarian Publications | SOUGAND in English