

Reconceptualizing the Target Domain of AI-assisted Writing: A Multi-Case Analysis of AI Literacy Integration in Writing Courses in Hong Kong Universities

Amy Kong*

The Hang Seng University of Hong Kong

Abstract

The application of Generative Artificial Intelligence (GenAI) in various writing contexts has transformed the writing process from a sheer manual cognitive task into human-AI collaborative writing. This paradigm shift necessitates a reconceptualization of writing constructs within higher education to ensure that writing assessments accurately reflect the evolving knowledge and skills demanded in the authentic AI-assisted writing context. Grounded in the argument-based validity framework, this qualitative study investigates the domain definition inference of three university English writing courses in Hong Kong. Data collection involved a comprehensive analysis of construct artifacts, including course outlines, teaching materials, and assessment task specifications, supplemented by semi-structured interviews with instructors. The data were mapped against Cardon et al. (2023)'s AI literacy framework comprising four dimensions: application, authenticity, accountability, and agency. Findings reveal significant discrepancies in operationalization of AI literacy across the faculties. All courses exhibited construct under-representation in the application dimension due to a lack of instruction on iterative prompting and operational mechanisms of Large Language Models (LLMs). One case even demonstrated construct irrelevance by utilizing AI detection reports that inhibit human-AI collaboration. New assessment specifications are proposed in the end to bridge the gap between conventional instructional/assessment design and the requirements for effective AI-mediated writing.

Keywords

GenAI-assisted writing, AI literacy, writing, argument-based validity framework

Introduction

The term *test* or *assessment* is often regarded as the process of obtaining inferences about language performance from elicited samples of language users' behavior when completing a language task (Chapelle & Lee, 2021). Various approaches to validating a language test or assessment have been developed to verify it does measure what it purports to measure. One distinct approach is argument-based validity (Chapelle et al.,

*Corresponding author amykong@hsu.edu.hk

Kong, A. (2026). Reconceptualizing the target domain of AI-assisted writing: A multi-case analysis of AI literacy integration in writing courses in Hong Kong universities. *Asian Journal of English Language Teaching*, 35(4), 51-79.
<https://doi.org/10.65961/AJELT-2026-4-004>

2008; Messick, 1989), which outlines the conventions of validating assessments. According to argument-based validity, six inferences should be warranted to demonstrate the usefulness of a test/assessment: target domain definition, evaluation, generalization, explanation, extrapolation, and utilization/consequences (Chapelle & Lee, 2021, p. 59).

The process of argument-based validation starts from domain definition, which refers to “the degree to which the [assessment] design reflects characteristics of language use tasks that examinees are likely to encounter in the academic domain” (Sawaki & Sinharay, 2013, p. 3). Domain definition establishes the target language use (TLU) domain, which is the authentic situation in which the students perform the assigned task. To justify this inference, the target language use domain should be analysed to identify key knowledge, skills, and abilities, and to link them with those included in the curricula and assessments (Chapelle et al., 2008; Kane, 2013; Kong, 2024; Park, 2021), so that warranty should ensure that the knowledge, skills, and abilities required are representative of those required in reality. In the context of AI-assisted writing, this means that if professional writing now involves collaborative intelligence and agentic workflows, the observed domain (the course curriculum and assessments) should be reconceptualized to include AI literacy; otherwise, the assessment could suffer from construct under-representation, failing to measure the very skills essential for success in the target domain situated in the age of AI.

In the realm of higher education, this domain definition is currently undergoing a radical change. The proliferation of Generative Artificial Intelligence (GenAI) has reshaped the educational landscape. Educators and students have increasingly applied GenAI in their work (Abdaljaleel et al., 2024; Ansari et al., 2023; Deng et al., 2026; Ibrahim et al., 2023; Teng, 2025; von Garrel & Mayer, 2023). In particular, the agentic workflows during human-AI collaborative writing action, which involve AI executing the tasks with human supervision, have complicated the definitions of the knowledge and/or skills required for effective AI-assisted writing (Cardon et al., 2023). This shift could constitute a new definition of the target real-world domain that is framed under the AI literacy because writing has now evolved from a purely individual cognitive exercise into a socio-technical process of collaborative intelligence. In this newly defined domain, the role of human writers has transformed from a primary creator to a supervisor who needs to establish specific goals and critically identify gaps for human intervention. As a result, the target domain must be reframed under the AI literacy construct (Cardon et al., 2023) that includes the abilities to plan, command, and identify constraints not only the final text, but the intermediate reasoning chains produced by GenAI.

The focus of this paper is to examine whether the writing courses and assessments of three universities reflect the AI literacy required in the AI age. This research is structured as a multiple case study that aims to analyse three different academic faculties to provide a comparative depth regarding the characteristics of AI-assisted writing. It also aims to examine whether these characteristics are able to provide evidence to support the domain definition inference, which posits that the educational constructs should reflect the characteristics of the authentic domain where AI-assisted

writing is prevalent. In university writing courses where assessment drives learning (assessment-based instruction), the instructional materials and assessment task specifications collectively define the observed domain, which is then compared against the target domain (AI literacy framework defined by Cardon et al., 2023). Therefore, analysing the construct artifacts, including the course outlines, teaching materials, and assessment specifications, could offer the primary evidence for whether the university curricula have successfully reconceptualized writing to include AI literacy, or whether it remains fossilized in manual traditions.

Literature Review

The New Target Domain of Effective Writing: Emergence of AI-assisted Writing

AI-assisted writing has become authentic writing for the here and now, and its influence is especially salient for effective communication. Many scholars and experts have suggested that writing tasks in the workplace has constituted the most significant use of GenAI by professionals (Cardon et al., 2023; Cardon et al., 2024). As forecasted by Davenport and Mittal (2022), “It may soon be standard practice, for example, for such systems to [...] provide first drafts of emails, letters, articles, computer programs, reports, blog posts, presentations, videos, and so forth”. Emerging survey and narrative reports have also suggested AI-assisted writing has become prevalent in the workplace. According to the study conducted by McKinsey in June 2023, approximately 62% of the existing work done by business professionals can be automated with AI (Chui et al., 2025). A survey of 621 business leaders in the United States reported that ChatGPT is being used for writing job descriptions (77%), drafting interview questions (66%), responding to job applicants (65%), summarizing meetings and corporate documents (52%), and generating task lists (45%) (Williams, 2023).

Due to the prevalence of AI-assisted writing in real-life contexts, the target domain of effective writing has also undergone a construct shift. Writing is no longer merely a cognitive process of humans translating thoughts into text, but it has evolved into a socio-technical process of collaborative intelligence (Nelson et al., 2025). The widespread adoption of agentic workflows in the workplace—where humans act as supervisors to AI agents—means that the ability to prompt, verify, and integrate AI outputs is now a core component of the writing construct (Singla et al., 2025). Consequently, if an instruction-based assessment strictly prohibits AI use, its domain definition inference may suffer from construct under-representation (Messick, 1989), as it fails to measure the collaborative skills within the human-AI hybrid writing context essential to the modern domain. Similarly, inappropriate monitoring of AI use could introduce construct irrelevance (Messick, 1989).

AI-assisted Writing: The Double-edged Sword

AI could improve productivity and quality. In an informal study that demonstrates a replicated workplace, it was shown that the consultants used AI to finish 12.2% more tasks on average, completed tasks 25.1% more quickly, and produced 40% higher quality results than those without (Mollick, 2023). Economist Erik Brynjolfsson also predicted a rise in productivity of up to 50% due to usage of GenAI in the workplace (Haan, 2023). One of the reasons for its increased productivity is of course, due to its amazing power in drafting and editing texts speedily. In addition, despite its imperfection, GenAI appears to possess relatively strong reasoning skills. Bang et al. (2023) found that GenAI was around 64 percent accurate in the 10 reasoning categories,

with optimal performance in deductive, inductive, and abductive reasoning and less effective in nontextual semantic and commonsense reasoning.

On the other hand, AI-assisted writing has many limitations. As Hyland (2026) conceptualizes, while current GenAI tools are highly effective in addressing surface-level language features such as lexical choice, grammar accuracy, and conciseness, they remain selectively reliable and often fail at deeper levels of writing, including rhetorical competence, disciplinary conventions, and argumentation where human intervention is irreplaceable. GenAI such as ChatGPT depends on an enormous data set of human language to work, but it does not always have access to the most updated information, and hence it may produce inaccurate and even biased information (Deng & Lin, 2023). While early models of GenAI like ChatGPT (GPT-3.5) relied on static, potentially outdated datasets, the current AI tools feature models optimized for specific tasks. For example, search-integrated models such as Gemini utilize real-time Google access to provide updated source citations, while reasoning-based models like DeepSeek could minimize logic errors. Despite these advancements, no model is entirely free from hallucinations. Moreover, other concerns related to AI usage including plagiarism, security, and fake news also prevail (Guo et al., 2023). If these issues are left unmonitored, it may not only compromise the accuracy of the work but also threaten the accountability of the writer, who is highly likely to take responsibility for the credibility of the final content.

The Need of Reskilling: Developing AI Literacy as Communicators

Human writers are expected to collaborate with GenAI on all forms of writing in the era of AI, but at the same time, they need to optimize its strengths and mitigate its weaknesses. As such, it is imperative for practitioners in education to reconceptualize the knowledge and skills needed for effective writing in the context of AI-assisted writing. In fact, employer surveys have also indicated a sharp decline in the importance employers are placing on strong written communication skills as a requirement for hiring, while the need for an ability to collaborate with GenAI to produce a quality written product is increasing (Cardon et al., 2023). In addition, AI also changes how professionals engage in writing; for example, they spend less time drafting and more time editing (Noy & Zhang, 2023). With these new knowledge and skills identified, education practitioners should bring GenAI into classrooms through instructions and assessments by introducing a move from writing to prove knowledge to writing to manage knowledge, hence requiring a restructuring of curriculum to support cognitive scaffolding (Zhang, 2025). In line with this pedagogical turn, research on human-AI collaborative writing underscores that effective integration relies heavily on hybrid intelligence, where the technological affordances of GenAI must be balanced with critical teacher engagement and feedback literacy (Deng et al., 2026).

Scholars and experts have urged for the development of AI literacy, which is defined as “a set of competencies that enables individuals to evaluate AI technologies critically; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace” (Long & Magerko, 2020, p. 2). Educational scholars (e.g., Getchell et al., 2022) have emphasized the need of developing theoretical frameworks to outline AI literacy from the perspective of effective communication. Since the current research is focused on effective communication in writing, it proposes

to adopt the AI literacy framework in the context of AI-assisted writing (Cardon et al., 2023) as the theoretical ground for the analysis.

Based on the qualitative and quantitative analysis of the findings of a survey with 343 communication instructors, Cardon et al. (2023) presented this framework of AI literacy that outlines four key abilities in the context of AI-assisted writing: *Application, authenticity, accountability, and agency* (p. 277-279):

- **Application** refers to an understanding of AI tools and how well they align with communication tasks. This encompasses not only the ability to navigate smoothly on the interface through effective prompting strategies but also understanding of the operational mechanisms behind different Large Language Models (LLMs) and hence their respective strengths and limitations (Cardon et al., 2023). What's more, effective promoting strategies should take into consideration prompt engineering competency (Kim et al., 2025), by which promoting is not only confined to single-shot AI interaction, but iterating prompting—refining AI prompts to achieve desired results (Hwang et al., 2023). Iterative interactions were noted to lead to more effective and efficient exchanges with AI and hence the AI output (Kim et al., 2025; Knoth et al., 2024; Lee et al., 2024; Ren & Wang, 2025). Crucially, Hyland (2026) highlights that the educational potential of AI interaction heavily hinges on prompt design; carefully structured, iterative prompts can elicit explanations and step-by-step guidance that foster deeper cognitive engagement, whereas poorly framed single-shot commands risk reducing AI into a tool for passive correction rather than long-term skill acquisition
- **Agency** means writers retaining control to make their own choices. In other words, GenAI should only serve as a complementary tool to enhance human knowledge and skills instead of making humans over-reliant on it. Writers should still be able to make decisions independently. AI-human collaboration in writing is playing a more important role in education. Zapata et al. (2025) suggested AI could serve as effective AI reviewer in the peer and AI review activity system. This system works like the conventional human peer review activity system (Kong, 2021), where the AI reviewer could provide scaffolds to the human writer (Kong & Bui, 2019), and the human writer could determine whether to accept or reject AI scaffolds depending on varying factors such as high self-efficacy in the subject matter (Kong & Teng, 2023). This ensures human writers still exercise *agency* over the writing task.
- **Authenticity** involves focusing on genuine communication and prioritizing the human element. This is related to communicators' skills of inserting their own voice, personality, and style in the AI-assisted writing, as well as maintaining audience-orientedness and sincerity of the final product. In fact, in the age of AI, while people are getting more comfortable with AI collaboration, tensions around authorship and creative identity still persist (Lin, 2025). Many students have also reported challenging balance between incorporating AI-generated ideas in writing and retaining their authentic voice (Wang, 2025). This tension is further complicated by students' language ideologies, as they often negotiate authority between idealized native-like norms embedded in AI outputs and their desire to maintain their original L1-influenced authorial identity (Li & Pan, 2024). This dilemma further underscores the need of instructing and assessing students' AI literacy in addressing *authenticity*.
- **Accountability** refers to communicators taking responsibility for the accuracy of AI-generated content and using GenAI in a fair and equitable manner. In fact,

hallucination rates of ChatGPT reached 91%, underscoring the need for careful scrutiny of AI-generated data (Adel & Alani, 2025). As such, human writers need to be able to verify the correctness of the AI output, to ensure the logic of the arguments, as well as to credit the sources of the information.

Research Gap and Research Questions

Recent policy addresses in Hong Kong (2024-2025) have explicitly called for the development of AI literacy frameworks, yet implementation at the school level remains varied (Legislative Council Secretariat, 2025). Similarly, while an increasing number of universities in Hong Kong have claimed to commit to renovating their curricula and pedagogies in the age of AI, it remains an underexplored area whether writing courses in universities in Hong Kong have effectively integrated instructions and assessments of AI-assisted writing based on Cardon et al. (2023)'s AI literacy framework for effective AI-assisted writing. In the context of this study, effective integration is defined as the conceptual alignment of instructional/assessment design with the four dimensions (4As) of the AI literacy framework: Application, Agency, Authenticity, and Accountability. Conversely, integration is regarded as ineffective when it results in construct under-representation—failing to teach essential AI-human collaborative skills. Please note that Cardon et al. (2023)'s AI literacy framework has been chosen as the foundation of the analysis because unlike other models that may focus on coding practice only and/or broader educational impacts (e.g., Churches (2008)'s Bloom's digital taxonomy), it was specifically developed for professional communication and writing based on a large-scale survey of 343 communication instructors, ensuring that the 4 As reflect the actual pedagogical concerns of the writing classroom for effective communication.

The current paper aims to bridge the gap by analysing the construct artefacts including the course outlines, instructional materials, and assessment details, as well as interviewing the instructors of three English writing courses from three universities to determine whether or in what ways their instructional/assessment design align with Cardon et al. (2023)'s AI literacy framework. Ultimately, the mapping could warrant the domain definition inference, aiming to evaluate whether their instructional and assessment designs align with the requirements for effective AI-assisted writing in the age of AI.

The domain definition is the first step in building a validity argument, as it asks whether the observed domain (what is taught/assessed) matches the target domain (abilities required in the real-world professional writing in the age of AI, and Cardon et al. (2023)'s AI literacy framework provides the specific constructs (Application, Agency, etc.) that define these abilities in the target domain. If the instructional and assessment design aligns with this framework, it provides the necessary evidence to warrant the domain definition inference, indicating that the course is measuring the required human-AI collaborative writing skills in the age of AI. To achieve this objective, the following research question is proposed:

1. To what extent does the instructional design (including learning outcomes, teaching materials, and assessment task specifications) of university English writing courses conceptually align with the AI literacy framework (Application, Authenticity, Accountability, and Agency), thereby warranting the domain definition inference for effective AI-assisted writing?

Research Methodology

Research Design

This study employs a qualitative research design focused on the analysis of construct artefacts (course outlines, assessment specifications, and teaching materials), supplemented by semi-structured interviews with the course instructors. The methodological approach is grounded in the argument-based validity framework (Chapelle et al., 2008). This research focuses exclusively on the domain definition inference. By examining the construct artefacts and by mapping them with the AI Literacy Framework, the current research determines whether the course constructs (what is taught and assessed) align with the abilities required in the AI digital age.

Research Context and Data Collection

The study situates its analysis within three university English writing courses (Faculty A, Faculty B, and Faculty C). All these academic faculties are focused on teaching English, and they come from three universities located in Hong Kong. This research is focused on analysing how AI-assisted writing has been integrated into three writing courses for effective communication for different purposes:

Faculty A: A Discipline-specific Writing course, which aims to enhance students' advanced literacy skills and deepen their critical understanding of the characteristics and purpose of academic discourse for writing the research essay.

Faculty B: A Foundational University English Course, of which the objective is to enhance students' foundational English proficiency in reading and writing, as well as introduce them the basic ideas of academic literacy for writing an academic essay.

Faculty C: An English Course for Effective Communication in Workplace Contexts, aiming to develop students' English communicative skills in workplace contexts through authentic communicative tasks. The data collection included the following data sets:

Construct Artefacts

- **Course Outlines:** Outlining the broader objectives of the courses.
- **Assessment Task Specifications:** These include assessment instructions and guidelines. In current research, the assessment tasks are analyzed not as measurement tools, but as operational definitions of the writing constructs, outlining what knowledge/skills are explicitly required or prohibited in the assessment tasks.
- **Teaching Materials:** Lecture slides and handouts used to teach the AI-assisted writing skills. The instructions and pedagogies are driven by the assessment design.

Instructor Interviews

Semi-structured interviews were conducted with an instructor of the respective courses to triangulate the artefact analysis and provide supplementary information regarding the instructional design driven by the writing assessments.

Data Analysis

Data were analysed based on the four components within the AI Literacy Framework: *Application*, *agency*, *authenticity*, and *accountability*, which also serve as the primary coding categories:

- **Application:** Understanding the operational mechanisms of the LLMs and hence strengths and weaknesses of various AI tools & effective prompting strategies.
- **Agency:** The ability of the writer to retain control over decision-making and avoid over-reliance on AI.
- **Authenticity:** Prioritizing human element, by maintaining, inserting, and/or infusing personal voice and style catering to the communicative needs.
- **Accountability:** Taking responsibility for the correctness and ethics of the writing produced with the assistance of AI.

To address the domain definition inference, the course objectives, assessment details, and teaching materials were coded to identify conceptual alignments with the four components of AI literacy. Please note that to ensure the reliability of the findings, the coding process involved two coders: The primary researcher and another university English teacher specializing in applied linguistics. The coders compared their categorizations of the artefacts, such as whether the instruction materials and assessment task aim to guide students to produce AI-assisted output that represented “supervisory agency” or “agency by restriction”, and whether personal voice would be maintained and/or infused based on the instructions and assessment task requirements. Any discrepancies in the classifications were resolved through discussion until 100% agreement was reached. The analysis sought to determine whether the course constructs included necessary skills for AI-assisted writing. The analysis focuses on identifying construct under-representation, where the instructional design disregards important AI-assisted writing knowledge/skills, and construct irrelevance, where the design focuses on abilities that are no longer relevant to the target domain in the age of AI. It should be well-noted that this study focuses on the *a priori* validity. In other words, it mainly analyses the instructional/assessment design rather than the student performance. However, within the scope of argument-based validity, analyzing the design architecture is a necessary prerequisite to establishing whether the assessment warrants the conclusion that students are prepared for effective communication in the real-life contexts.

Findings

This section reports the findings from the document analysis of the construct artefacts (i.e., course outlines, teaching materials, and assessment specifications), as well as instructor interviews across the three academic faculties. Please note that the details of the assessment tasks, AI usage policy, and verification mechanisms are documented in the Appendix that is attached to the end of this manuscript. The findings are categorized by the four dimensions of the AI literacy framework: *Application*, *agency*, *authenticity*, and *accountability*.

Academic Faculty A: A Discipline-specific Essay Writing Course

In Faculty A, AI-assisted writing is adopted in a discipline-specific essay writing course. The objective of the writing course is to enhance students’ advanced literacy skills and deepen their critical understanding of the characteristics and purpose of academic discourse for writing the research essay.

The assessment instructions explicitly allow students to write the essay with the assistance of AI, but the assistance is only limited to brainstorming ideas for writing and improving minor language issues such as stylistic, correcting grammatical mistakes, or looking for appropriate synonymous expressions. Students have to submit all AI interaction logs, along with their draft and final essay, and attend a five-minute viva upon submitting the final draft.

Application (AI Tool Knowledge & Information Retrieval)

According to the teaching materials, students are asked to brainstorm various GenAI tools that would be useful for writing the research paper. However, no comparisons or deep analysis of the tools is found in all the artefacts. In terms of prompt engineering, simple guidance and samples on extracting key information from the given reading sources are also provided. The guidelines instruct students to use AI to locate information, but no specific instructions were identified regarding effective prompt writing and iterative prompting techniques for evaluating the reading sources or synthesising main arguments from multiple sources, nor any instructions are found on prompting AI to brainstorm ideas effectively.

Authenticity & Agency (Restricted Use of AI)

In accordance with the assessment requirements, students are granted permission to seek AI assistance only for brainstorming ideas and improving minor language issues. No instructional materials were found that explicitly teach skills for integrating personal voice or maintaining individual stylistic uniqueness into the AI generated output, as the writing is expected to be manual. Since the outlines clearly restrict students' use of AI, it appears to address *agency* because students are required to write majority of the content manually and independently, though this definition is on the surface level.

Accountability (Monitoring)

Accountability is also addressed because students are required to submit all AI interaction logs, along with their draft and final writing. In addition, students have to attend a five-minute viva after submitting their final paper. According to the instructor, students are asked to clarify certain parts and respond to questions about the changes made between drafts and the reasons for using specific references during the viva.

Academic Faculty B: A Foundational University English Course

Faculty B tries to integrate AI literacy into a University English course compulsory for all first-year students. The course aims to enhance students' foundational English proficiency in reading and writing, as well as introduce them the basic ideas of academic literacy for writing an academic essay. The course is also designed to introduce AI-assisted writing in the writing process, including the ethical uses of AI, effective skills for generating AI prompts, and evaluation of AI feedback.

As for the assessment, students need to consult AI for its feedback on their essay draft. They have to evaluate the AI feedback and determine whether or how to integrate the feedback in their final draft. Students need to submit all the interaction logs, as well as a reflection outlining how AI tools have assisted them in modifying the essay, along

with their first and final essay draft. The reflection accounts for 5% of the total weighting of the whole essay writing assessment.

Application (Basic Introduction to Prompt Writing)

The teaching materials in Faculty B contain explicit instructions on basic prompt writing. Students are demonstrated how to create effective prompts to enhance the quality of AI output. For example, they need to define the context and role for the writing paper (e.g., Imagine you are an English student writing an argumentative essay on...). Students are also instructed to be specific when outlining the task they want AI to perform, and to provide relevant details such as output format and background information about the task. For example, students may ask AI to specifically comment on the strengths and weaknesses of the counter-rebuttal they have written in the essay based on the attached reading sources. However, the materials do not include guidelines or examples about when and how to generate iterative prompts to refine AI output.

Accountability (Critical Evaluation of AI Feedback)

After students ask the AI to give feedback to their draft, students are instructed to evaluate the feedback before integrating it in their draft. There are guiding questions on making effective evaluation, but they seem to be guiding students for prompting the AI to give feedback to their essay draft (e.g., Is the thesis statement given? Does the paragraph have the topic sentence? Is academic language style used?), instead of ones for evaluation the AI feedback. The teaching materials did illustrate the importance of logical flow of ideas in essay writing based on the argument-evidence-explanation model, followed by presentation of an instance of logical jump in a manual essay writing, and that AI feedback has missed spotting this logical jump. However, what constitutes AI missing of logical jumps, and what follow-up prompts should be generated to refine or further elicit AI suggestions, are missing in the instructions.

Agency & Authenticity (Critical Selection of AI Suggestions)

The instructional design promotes *agency* and *authenticity* through critical selection of AI suggestions instead of complete adoption. The teaching materials include samples illustrating how students might accept, reject, or modify specific sentences generated by the AI. For example, with the instruction on the argument-evidence-explanation model, students are empowered to identify the logic jumps in the AI feedback. This ensures students' assertion of their unique ownership of logic.

In another teaching sample, students are shown to reject an AI suggestion that does not fit their communicative goals (e.g., declining the AI's suggestion to add statistics as this would make the tone too technical for a general academic essay writing). This is another clear instance showing how students could preserve their own authorial style over the AI suggestion, addressing both *authenticity* and *agency*.

While the assessment-based instructional materials and assessment briefs seem to encourage critical and reflective AI integration, the interview data reveals a gap in implementation. The instructor reported that majority of students relied on single-shot prompting (e.g., asking AI to comment on the thesis statement) and accepted the AI revisions completely without critical analysis or follow-up clarification. This complete uptake of AI suggestions seems to be a kind of plagiarism in the instructor's opinion. Furthermore, perhaps due to the low stakes of the students' reflection (weighted at only 5%), almost all students submitted generic reflections lacking in critical depth. The

instructor also admitted that due to the absence of specific assessment criteria for the interaction logs and high workload pressures, she rarely scrutinized the logs during grading. This suggests that the assessment design may fail to enforce the intended “Supervisory Agency” in practice.

Academic Faculty C: Effective Communication in Workplace Contexts

Faculty C adopts AI-assisted writing in a compulsory English course for senior students, aiming to develop students’ English communicative skills in workplace contexts through communicative tasks. An assessment requires students to write a personal narrative for applying a specific job position. Students are coached step-by-step on how to utilize AI to generate the draft.

However, despite the clear demonstration of the templated prompts for AI to generate the draft, students have to be cautious of the extent they integrate the AI-generated ideas, as marks will be deducted if the AI detection reports exceed a pre-set limit. The assessment descriptors do not include any features of AI-assisted writing, but include the assessment of “original ideas”. The instructor admitted that any suspected AI-assisted phrasing lacking personal evidence could result in mark deductions.

Application (Template-based Prompting)

Adopting a template-based approach, instructors provide students with sample prompts for asking AI to generate drafts for the personal narrative. First, they prompt the AI to identify the main skills and qualities that employers seek for a specific position listed in a job ad. Then, students need to provide the AI with the required skills/qualities, together with their individual qualifications, and ask it to suggest an outline for a 4-paragraph personal narrative with keywords, detailed idea, and examples. Despite this step-by-step demonstration, the instructor reported that students’ uptake of the AI-generated content was sometimes limited, as some students chose to reject the ideas because they did not match their desired narrative move, but exactly how many students retained such authorship remains unknown to the instructor because no drafts were collected. .

Accountability (Surface Evaluation & Policing)

According to the teaching materials, students are asked to evaluate the outline generated by AI. For example, the GenAI tool in the sample materials has produced 5 paragraphs, the students are advised to consolidate it by deleting one qualification and merging another. Such evaluation samples appear to focus on surface-level instead of deep analysis of the content. Nonetheless, according to the instructor, students need to adhere to the “original ideas” criterion, and marks are deducted if the rate in the AI detection reports exceed a pre-set limit agreed by the teaching panel.

Agency (Policing)

According to the instructor, students need to write the assignment according to the rubrics, which includes a component called “original ideas”. As a result, some students might disregard the AI-generated content that is not specific enough to illustrate their specific attributes, skills, and experiences. For example, some students may choose not to emphasize the phrase “being passionate about customer services” because they are not genuinely passionate about the job; rather they are applying simply due to necessity. This would allow more control of the writing process and hence maintain agency.

However, again, since no draft is collected, the instructor has no idea how many students may reject AI ideas.

Authenticity (Traditional Narrative Instruction)

Since this is a personal narrative, the assignment actually accentuates importance to personal voice. In fact, instructional materials are presented to address the ideas of persona. Teachers guide students to analyse a sample writing from a previous student to demonstrate how to narrate a unique story that could establish one's professional persona. This instruction aims to help students address the "original ideas" marking criterion. However, the analysis of the sample writing seems to be a sheer demonstration of a manual writing sample, instead of instructions that coach students how to use AI to refine the writing style (e.g., prompting AI to adopt a specific persona), or how to integrate one's own unique authorship with AI-generated ideas. In addition, no specific teaching materials that explicitly coach students how to establish their persona through various appropriate tones or style were observed.

Discussion

The analysis of the construct artefacts and instructors' interview data across Faculties A, B, and C reveals a diverse approach to integration of AI literacy into AI-assisted writing assessments. While all faculties acknowledge the necessity of co-authoring with GenAI in the digital era, their implementation of the 4 As (*application*, *agency*, *authenticity*, and *agency*) varies significantly, ranging from restriction (Faculty A) to penalty (Faculty C), with only Faculty B demonstrating a certain degree of assessment of students' ability to critically integrate AI-generated ideas in their writing.

Application: From templated prompts and Simplistic Extraction to Cognitive Scaffolding

The *application* component of the AI literacy framework refers to the practical, operational ability to use GenAI tools effectively to achieve specific communicative goals (Cardon et al, 2023). The findings highlight a stark difference in how *application* is conceptualized taught, and assessed across the three faculties, revealing a move from templated prompting and simplistic extraction to more complicated cognitive scaffolding.

Faculty C adopts a template-based approach, providing students with sample prompts to ask AI to generate drafts for their personal narrative. On the other hand, Faculty A instructs and permits students to use AI for writing, but simply for extraction of ideas from reading sources. These pedagogical choices treat AI as a tool for drafting and summarization rather than a co-writing partner. They fail to encourage students to collaborate with LLMs for generating the output, hence sidestepping students' mastery of the complex AI literacy skills for managing AI-generated texts.

In fact, the purely template-driven prompts employed by Faculty C and simple commands for information extraction suggested by Faculty A are similar to the student-directed prompts in Kim et al. (2025). First, students simply ended their engagements with AI upon receiving answers, leading to almost nil turn-taking dynamics. In addition, their prompts demonstrated a recurrent pattern. These kinds of static and recurring prompt patterns often undermine student-AI interactions rather than engaging them in

active participation and even reasoning process necessary for skill acquisition in the age of AI.

In contrast, Faculty B operationalizes *application* as effective promoting through cognitive scaffolds. By focusing on context setting and role definition, as well as the details of the task requirements, Faculty B treats prompting as effective communication skills rather than a quick technical fix. This strategy partially echoes with the prompt patterns reported in Kim et al. (2025)' study, where "high literacy students exhibited descriptive and context-based prompts..." (p. 1). These high literacy students also demonstrated the ability to use varying verbs depending on their writing need. In addition, the inclusion of the context setting and role definition in the prompt also reflect the use of persona-based prompts, which identify the voice of the writer in thinking through audience and tone (Rice, 2025).

However, two critical gaps exist across all faculties: First, both the assessment and instruction overlook students' knowledge of the strengths and weaknesses of various GenAI tools. For example, different models (e.g., GPT-4, Perplexity, Claude, DeepSeek, Gemini) have distinct architectures optimized for different tasks (e.g., creative writing, deep research, information retrieval, fact-checking, & synthesizing). Without this knowledge, students may simply prefer the most popular tool rather than the most appropriate one, leading to undesirable outcomes. In fact, LLMs often generate clustered responses with similar meanings after being prompted multiple times (Wen & Hu, 2025). This consistency can make them seem reliable, but without being aware of the operational principles of the models, users may underestimate the risks of over-trusting the responses and stray away from human reasoning.

Second, the instruction and assessment criteria miss the requirements for iterative prompting (Kim et al., 2025; Ren & Wang, 2025) across the three courses being studied. For example, Faculty A only suggested students could summarise the main ideas of the given reading sources with the AI tool, and Faculty C even instructed the students to ask AI to generate the whole draft using templated prompts. Both faculties did not require students to submit the interaction logs and so no assessment of the prompt quality is needed. Faculty B did teach students the prompting skills but there was no emphasis on iterative prompting. Even though students were asked to submit the interaction logs and reflection, the small percentage of weighting of the reflection (5%) and no specific requirements for quality prompting have made teachers simply gave impressionist marking to students based on the reflection, without scrutinizing the prompting process. In fact, according to the interview with the instructor, most students simply asked AI based on one prompt only. For example, they would simply ask AI to check the quality of the thesis statement and made revisions according to AI's edited version. This could limit students to merely single shot interactions (see the example below), rather than engaging in fruitful interactions with AI through iterating prompting.

Single-shot Prompting (The most frequent approach adopted by the students in Faculty B, according to the interview with the teacher):

Prompt: Can you comment on my thesis statement?

AI: It is clear, but a better revision is... (Students simply revised the draft based on AI's suggestion).

Example of Iterative Prompting (But this is missing in the students' works according to the interview with the instructor, probably due to no requirements for rating the interaction logs):

Prompt 1: Evaluate my essay draft and provide feedback on the logical flow of my arguments

AI Response: There is a logic jump between Point A and B. A better revision is....

Iterative Prompt: However, I think that the example is already bridging Point A and Point B. Why do you think this is still a logic jump?

AI Response: This is because....

Iterative Prompt: Then can I revise the writing in this following way...

Iterative prompting is vital because it transforms AI from a static answer machine into a collaborative partner, so that both sides could revise the idea to produce an even more complex one through reflective dialogues, and that students could demonstrate their critical thinking skills and refine their understanding, whereas a single-shot interaction often leads to surface-level content or sometimes even plagiarism (since students could simply copy the AI's suggestions). In fact, Teng (2026) also highlighted the importance of strategic engagement and supporting a reconceptualization of vocabulary learning as a dynamic human–AI interactional process. Without training in such a dynamic AI-human iteration, students are less likely to develop advanced AI literacy required for co-creation of more complicated written artefacts. Iterative prompting was reported to be a common strategy employed by the high literacy students in Kim et al. (2025)'s study. These students created prompts interactively, encouraging collaborative information exchange through cohesive and reflective dialogues with AI. This approach enables them to evaluate or critique AI responses, refine their understanding, and keep iterating, allowing ongoing creative discussions and enhancing quality content generation collaboratively.

Agency: The Tension between Restriction and Empowerment

Agency in the context of AI literacy is defined as the user's ability to retain control over the decision-making and production process, making sure that AI simply plays an assistant role instead of a substitute of humans (Cardon et al., 2023). The comparative analysis illustrates diverse approaches to how *agency* is addressed across the three faculties, divided between agency by restriction and agency by supervision.

Both Faculties A and C seem to define *agency* as writing the task without AI. Faculty A attempts to preserve students' agency by restricting AI use to brainstorming ideas or editing minor language issues in the final assessment, and by checking whether students conform to this restriction through the viva. Faculty C deters students from using AI in their actual assignment by enforcing the "original ideas" rubrics and AI detection reports. This *agency by restriction* approach creates a false dichotomy between human writing and AI writing, and it frames *agency* as the absence of AI, rather than the mastery of it. This approach deviates from the most updated Adaptive Human–AI Writing Continuum (AHAWC) Model, which integrates five collaboration patterns (Human–AI–Human, AI–Human–AI, Human-in-the-Loop, AI Companion, Human–AI Co-creation (Estacio, 2025), where hybrid authorship is emphasized for effective AI-assisted writing.

In contrast to maintaining agency by restriction like Faculties A and C, Faculty B offers a more sophisticated approach in which agency is defined as overriding AI. For example, the instructional materials demonstrating the logic jumps identified by students but overlooked by AI is a vivid manifestation of how agency could be best cultivated by positioning the student as the final decision maker, rather than simply banning the tool. This approach resonates profoundly with the concept of “Human-in-the-Loop (HITL)” in agentic workflows (Johnson et al., 2025, p. 35). HITL ensures that human actors remain central, especially when making decisions require contextual judgement and reasoning. By teaching students to identify where the AI fails (e.g., logic jumps) and assessing such ability in the final reflective note, Faculty B fosters *supervisory agency*. The student acts like a supervisor of AI, where *agency* is expressed through the power of rejection and modification to AI-generated ideas. Faculty B’s emphasis on the student’s power to override the AI suggestions could ensure that they retain ownership of the voice and the writing product. Moreover, using ChatGPT for feedback in writing was reported to impose a positive effect on students’ writing motivation, self-efficacy, engagement, and collaborative writing tendency (Teng, 2024), and so Faculty’s B instructional design seems to aim not only at retaining agency, but even promoting the writing process.

However, the interview data reveals a construct gap between this sophisticated intent and the actual assessment outcome. Since the assessment task specifications do not specify how to produce dynamic interactions (e.g., iterative prompting) and partly due to the low weighting of the reflection (5%), almost all students demonstrated only shallow, single-shot interactions with AI. For example, some simply asked AI to generate an outline or limited themselves to the sheer integration of suggestions. This indicates that while the teaching materials encourage agency, the ambiguous criteria and low stakes of the assessment specifications act as an inhibiting factor for engaging students in true HITL workflows (Johnson et al., 2025). Nonetheless, it should be noted that the current research focuses on warranting only the domain definition inference (i.e., analysing the design); no analysis of students’ writing was involved. Future studies should be conducted that warrant the evaluation, explanation, and/or extrapolation inference (See Dursun & Li, 2021 for details), which involves analysing students’ writing drafts to demonstrate that the assessment task elicits the desired learning outcomes that are important for students to perform effectively in the AI-assisted writing contexts.

Authenticity: From Avoidance, Paradoxical Pedagogy, to Supervisory Agency

Authenticity refers preservation of authorship by keeping the author’s voice and intent (Cardon et al., 2023). Many studies have reported the dilemma writers could face in generative AI-assisted writing: How to maintain balance between incorporating AI output in writing and maintaining their authentic voice (Cardon et al., 2023; Lin, 2025; Wang, 2025). Many writing instructors also expressed concerns over the loss of voice, which may result in “writing all sounds alike” (Cardon et al., 2023, p. 44).

The findings of the current study also reveal three distinct avenues of operationalizing *authenticity*: Avoidance (Faculty A), paradoxical pedagogy (Faculty C), and critical selection for AI feedback (Faculty B). These differences illustrate the complexity of defining authorship in the AI age.

Addressing *authenticity by avoidance*, Faculty A attempts to protect *authenticity* by restricting AI utility strictly to simple peripheral tasks like brainstorming and minor language polishing. While this effectively forces manual writing that keeps human

authorship, it fails to prepare students for contemporary writing contexts where human-AI hybrid authorship is prevalent, and hence students could be ill-equipped to maintain their voice. Worse still, Faculty C even presents a pedagogical paradox in terms of *authenticity*. The goal of the course is to prepare students for effective communication in workplace contexts where persona is of paramount importance. However, the findings reveal a disconnect with the pedagogy, where AI is simply used for creating drafts while establishment of persona through voice is introduced through manual analysis of previous student samples.

Notably, both Faculties A and C have missed the opportunity to teach and assess AI-assisted authenticity—coaching students on how to prompt AI to adopt a specific tone (e.g., “Write in a passionate, professional tone...”) or how to infuse a generic AI-generated draft with specific contextual judgements and/or anecdotal evidence. In other words, *authenticity* has been treated as a trait that exists outside the spectrum of AI interaction, rather than a skill to be handled within it. The lack of coaching of strategic handling of one’s unique voice within the AI interaction spectrum once again hampers students’ development of human-AI hybrid authorship. If instructors of writing courses continue to avoid confronting the dilemma of balancing AI-generated ideas and human authorship, students could never learn how to edit and integrate AI responses in their own unique writing style that addresses the context and target audience. This underscores the necessity of a complementary approach advocated by Hyland (2026), who argues that the most effective pedagogy does not position AI as a substitute for human evaluation, but as a complement that absorbs routine corrections, thereby freeing up teachers’ expertise to facilitate the complex, meaningful, and relational dimensions of student writing.

In contrast, Faculty B offers the most desirable prototype, closely aligned with the AI literacy framework, as *Authenticity* is defined not by the avoidance of AI, but by students’ rigorous supervisory agency. Instructional materials demonstrate students actively overriding the AI feedback. For instance, rejecting a suggestion to add statistics because it would make the tone too technical for a general academic essay writing. Such ability is further assessed in the reflective note as well. This representation of authenticity through *supervisory agency* acknowledges the fact that an AI-generated text can also become authentic as long as the student has the ability to exercise their power to accept, reject, or modify its output based on their communicative intent.

Accountability: Policing vs. Verification

Accountability refers to writers taking responsibility for the accuracy and credibility of the final writing created with the assistance of AI (Cardon et al., 2023). The three faculties demonstrate a significant divergent approach between policing and verification, highlighting controversies in the instructional and assessment design.

The reliance on the AI detection reports and pre-set mark deduction scales instils a compliance mindset in students of Faculty C. As a result, students may simply verify their work against a probability score determined by the AI detection tool, not against truth. What’s worse, *accountability* here is externalized to the machine, which could be

unreliable and inconsistent (Adam et al., 2025; Han et al., 2025). If the detector shows a “pass”, the student is considered accountable, regardless of the credibility of the work.

On the other hand, both Faculties A and C enforce *accountability* as verification. Faculty A exercises verification through a high-stake viva, where the student has to demonstrate possession of the knowledge in the essay they write. This moves beyond policing plagiarism to verifying the student’s understanding. Moreover, Faculty B even enacts verification through coaching students on identifying and fixing AI logic errors, followed by outlining this ability in the reflection note. This fosters a deeper form of critical intellectual responsibility than simply ensuring a detection score is low. Both Faculties A and C align advice to move towards adopting multifaceted approaches to evaluation to ensure academic integrity (Han et al., 2025). Most importantly, by enforcing students to explain the reasons for their critical selection of AI-generated ideas through either oral defence and/or reflection notes, they are also training students to make evaluative judgement of the AI outputs, a unique human capability imperative at a time of technological explosion (Bearman, 2024).

Warranting the Domain Definition Inference: Validity Threats

In summary, according to the comparative analysis of the construct artefacts of the writing courses across three faculties, the instruction-based assessments struggle to fully warrant the domain definition inference. The journey toward valid AI-assisted writing assessment remains challenging not only due to a lack of adoption, but also because of the flaws in pedagogical and assessment designs. For example, the findings highlight difficulty in instructors’ changing the mindset from policing to embracing the concept of HITL agentic workflows, as well as a misalignment between instructional/assessment goals (what is taught and assessed) and assessment specifications (what is graded). These barriers could result in notable gaps within the four As (Application, Agency, Authenticity, and Accountability) within Cardon et al (2023)’s AI Literacy Framework in the AI-assisted writing context (See Table 1):

Application: All faculties display under-representation in the *application* dimension because they fail to teach and assess students’ knowledge of the operational mechanisms of LLMs and their skills of iterative prompting. Students are limited to static interactions, which does not reflect the complex prompt patterns for effective AI interactions. However, Faculty B demonstrates a better alignment with the framework by providing cognitive scaffolds (e.g., context setting, role definition, and communicative needs), while Faculties A and C simply teach basic extraction or templated prompts that fail to equip students with effective prompting skills. That said, Faculty B still fails to assess the quality of the prompts. Without clear specifications of the assessment criteria for effective iteration, students lack the incentive to engage in the complex, dynamic dialogues required for effective AI interaction.

Agency and Authenticity: Faculties A and C exhibit construct under-representation by defining these constructs through the absence of AI. They fail to represent the human–AI co-creation of writing prevalent in the target domain. In contrast, Faculty B better aligns with the domain definition by operationalizing supervisory agency in the instructional materials. Both agency and authenticity are addressed by empowering the students to critically select, reject, or modify AI suggestions, reflecting the *Human-in-the-Loop* (HITL) workflows for effective AI-assisted writing. However, the interview

data uncovers a critical implementation challenge: Assessment-Instruction misalignment. Although Faculty B's teaching materials encouraged the agentic HITL workflows, the assessment specifications, such as the low weighting (5%) and nil criteria for the logs and reflection, resulted in students performing only shallow, single-shot prompting. This demonstrates that "warranting the inference" is difficult because it requires not just updated instructions and assessment briefs, but a fundamental restructuring of grading mechanisms such as increased weighting and clear specifications to reward the mastery of AI.

Accountability: While Faculties A and B warrant the accountability construct through verification (viva and reflective notes), Faculty C demonstrates construct irrelevance. By relying on AI detection reports, Faculty C seems to shift the assessment focus to the student's ability to mask AI traces rather than their ability to verify the accuracy or credibility of the AI-generated content. The assessment of this ability is irrelevant to the communicative goals of the target domain in the age of AI, where accountability is defined by the accuracy and ethics of the content, not the probability score of a detection algorithm.

Table 1. Mapping of Instructional Designs to AI Literacy Constructs

AI Literacy Dimension	Faculty A	Faculty B	Faculty C	Construct Validity Status
Application: Tool Knowledge	✗ No analysis of model strengths/limitations.	✗ Knowledge of distinct tools is missing	✗ Knowledge of distinct is missing.	Construct Under-representation: All faculties fail to teach model-specific logic.
Application: Prompting Strategies	Δ Limited to basic extraction commands	Δ Cognitive Scaffolding for writing prompts, but no specific instructions on iterative prompting skills.	Δ Static, template-based on prompting.	Construct Under-representation: Lack of iterative prompting across all courses.
Agency: Independent Decision-Making	✗ Restricted use; agency via avoidance.	Δ Supervisory agency; empowered to override AI in the instruction materials, but a gap in the assessment specifications.	✗ Policing mindset; agency in AI-prohibition.	Construct Under-representation (A & C): Agency defined as minimizing or prohibiting AI, not mastery of it. (B): Misalignment between instructions and assessment specifications.
Authenticity: Human Voice & Persona	✗ Manual writing expected; Voice exists outside AI	✓ Authenticity through critical selection of AI feedback.	✗ Paradoxical; AI used for drafts but manual voice required in assessment.	Construct Under-representation (A & C): Failure to teach AI-assisted authenticity.
Accountability: Accuracy & Ethics	✓ Verification via high-stakes viva.	✓ Critical logic checks & reflection notes.	✗ Compliance score via detection reports.	Construct Irrelevance (C): Evaluates masking of AI rather than truth/ethics.

Proposed Specifications of AI-assisted Writing Assessments

Based on the research findings, this research proposes to integrate the following specifications in the AI-assisted writing assessments, which should further transform the assessment-driven instructional design and even learning outcomes of the curriculum of university writing courses for effective communication curriculum in the age of AI:

Application: Beyond Basic Prompting Skills

- **Operational Mechanism Knowledge:** Assessments should require students to show understanding of the architectural design or at least the optimized tasks (e.g., creative vs. factual) of existing LLMs to ensure selecting appropriate tools to cater for the needs of their assessments.
- **Cognitive Scaffolding:** Assessments are expected to require students to generate context-based, persona-oriented prompts (e.g., defining roles, contexts, & communicative needs) rather than simply extraction of ideas or passive uptake of static templates.
- **Iterative Prompting History:** Assessments should include students' documentation of a chain of reflective dialogues with AI that demonstrate collaborative information exchange and refinement instead of simply a single-shot command.

Agency: Mastery of Supervisory Competency

- **Human-in-the-Loop (HITL):** Assessment briefs should redefine agency as the mastery of the AI tools through HITL workflows instead of mere restriction or absence of AI.
- **Evidence and Justification of Supervisory Role:** Assessments should require students to submit documents that show instances where they override AI suggestions to meet communicative needs. Students should also submit AI interaction logs and reflective notes that justify the reasons behind their decisions throughout the human-AI collaborative writing process.
- **Assessment Specifications:** The assessment of the HITL integration should carry significant weight (e.g., 20-30%) to incentivize students' engagement in agentic workflows. In addition, marks are awarded for the quality of the student's *supervision*—specifically their ability to identify logic gaps or inconsistencies in the AI output.

Authenticity: AI-Assisted Human Style/Voice

- **Defining Persona and Strategic Infusion with Uniqueness:** Assessment-based instructions should guide students how to develop their persona that address specific communicative and audience needs. Assessments should take into account students' ability to infuse AI-generated content with unique individual style/voice and anecdotal evidence through strategic prompting.

Accountability: Intellectual Responsibility

- **Fact and Logic Validation:** Assessment outlines should draw students' attention to their ability to identify and rectify hallucinations or logic jumps generated by AI.
- **Human-centred Verification:** Assessments should include multi-modal verifications such as oral defence and/or documentation of AI interaction logs

to authenticate students' cognitive ownership of writing, instead of relying on sheer unreliable AI detection scores as a measure of integrity.

Based on the above proposed specifications, a proposed assessment rubric for assessing the AI-human collaborative competence is drafted below (Table 2).

It should be noted that integrating these specifications in the assessment criteria could be pivotal in the context of English for Effective Communication, especially in the English for Academic Purposes (EAP), where the primary goal is not simply to produce text linguistically accurately, but to develop one's critical reasoning and express their ideas catering to the communicative goals. This adoption of the proposed rubric could transform AI-assisted writing from a potential threat to academic honesty into a demonstration of one's evaluative judgment and infusion of personal voice based on the communicative contexts, the skills that are essential for navigating the complex information landscape of higher education.

Table 2. Proposed Assessment Rubric: The AI-Human Collaborative Competence

Dimension	A Range	C Range	F Range
Application	• Demonstrates a sophisticated prompting strategy that involves substantial cognitive scaffolds (e.g., role and context definitions). • Evidence of iterative prompting, where the student actively guides the AI to improve output quality. • Justifies the selection of specific tools based on the strengths of individual models.	• Prompts are clear but sometimes lack effective cognitive scaffolds (e.g., missing role definitions or context). • Uses effective but largely single-shot prompts. • Tool selection is appropriate but generic.	• Relies only on single-shot commands (e.g., “Write an essay about X”). • Uses inappropriate tools for the task (e.g., using a creative LLM for retrieving facts without verification).
Agency	• Student explicitly identifies and rejects/modifies AI suggestions that do not meet communicative goals. • Reflection justifies why specific AI ideas were rejected based on communicative needs or logic.	• Student retains control over the main thesis but passively accepts AI’s underlying argument structure or logical flow. . • Some evidence of critical selection but occasionally defers to AI logic without deep questioning.	• Either avoids AI entirely (manual writing only) OR accepts AI output entirely. • No evidence of human modification or critical judgment of AI suggested ideas, with AI dictating the argument structure.
Authenticity	• Successfully infuses a unique persona or specific local context that the AI could not generate alone. (e.g., substituting the AI’s vague expression “I have teamwork skills” with a detailed narrative about “having the experience of resolving	• The writing is polished and professional but retains a detectable AI voice/style. (e.g., overusing unnatural transition words or flowery AI phrases like “unwavering support” that do not match the student's natural persona).	• The writing lacks a distinct human voice, but heavily relies on generic AI tone (e.g., adopting a highly detached, academic tone suggested by AI for a personal job application, showing no human adaptation).

	- a group project in the XXX Competition”. - The final voice is consistent, professional, and distinct from the generic AI tone.	Personal voice is noted but occasionally overshadowed by the AI syntax	The writing fails to adapt the content to the specific target audience.
Accountability	- Student successfully identifies and corrects hallucinations in the AI draft. - All claims are verified against primary sources. - Full transparency in the AI declaration regarding which parts were AI-generated and which were human-written.	- Basic verification of major facts, but may miss subtle hallucinations or logic jumps. - Citations are generally accurate and AI use is declared but lacks specific detail on the extent of assistance.	- The writing contains unverified hallucinations (e.g., fake citations or false facts). - The writer blindly trusts AI logic without human validation. - The writer fails to declare AI use.

Conclusion, Implications, and Limitations

By analysing the strengths and weaknesses of the construct artefacts across three university writing courses that involve AI-assisted writing, the current paper has identified critical gaps of the existing instruction and assessment design compared with the four components (Application, Agency, Authenticity, and Accountability) within the AI literacy framework. These discrepancies contaminate the validity of the assessments, especially in warranting the target domain inference.

The proposed specifications of AI-assisted writing serve as a new guidepost for education practitioners to transform and refine their existing assessment and instructional design, ensuring what they teach and assess reflect the knowledge and skills required for Human-AI collaborative writing, which goes pervasive in both authentic academic and professional contexts in the age of AI.

Despite practical insights, the current research primarily examined the target domain inference, other facets of the argument-based validity framework, for example, evaluation inference (how the assessment process and rating rubrics elicit behaviours), generalization inference (measuring inter-rater reliability), explanation/extrapolation process (linking performance to real-world writing output), and utilization/consequence inference (users' perspectives) are not covered. More research studies should be conducted to analyse both qualitatively and quantitatively regarding the validity of the assessments adopted in the AI-assisted writing curriculum. For example, the students' drafts could be validated to warrant the evaluation, explanation, and extrapolation inferences. Interviews with the students could also be conducted to explore the washback effects of the proposed assessment rubric on their learning process.

Appendix: Anonymized Assessment Task Specifications

Note: To preserve the anonymity of the three institutions, the following details are revised information that outline the assessment design, AI usage policies, and verification mechanism noted in the three cases.

Faculty A (Discipline-Specific Academic Writing)

- **Course Context:** Advanced academic literacy for a specific discipline.
- **Assessment Task:** A research-based academic essay
- **Instructional Focus:** Ask students to brainstorm various AI tools.
- **AI Usage Policy:** Generative AI is allowed only for brainstorming and minor stylistic or minor language-related matters (e.g., synonyms).
- **Verification Mechanism:** A mandatory viva (50% weighting of the final essay) is conducted where students must answer content-related questions to demonstrate their understanding of the argument.

Case 2: Faculty B (Foundational University English Writing)

- **Course Context:** General English proficiency
- **Assessment Task:** An argumentative essay
- **Instructional Focus:** Lectures demonstrate how to write AI prompts with context/role definitions and identify logic jumps in AI output.
- **AI Usage Policy:** Students are encouraged to use AI to critique their drafts and override weak AI suggestions (e.g., logic jumps).
- **Verification Mechanism:** Students need to submit a reflective note on AI Use (5%), together with the AI interaction logs.

Case 3: Faculty C (English for Workplace Communication)

- **Course Context:** English for professional workplace communication
- **Assessment Task:** A personal narrative for a specific job posting.
- **Instructional Focus:** Templated AI prompts for generating the outline.
- **AI Usage Policy:** While AI is used for generating the outline, the final submission must show students' original ideas, including demonstration of unique voice and personality.
- **Verification:** Submissions are run through AI detection software. Marks are deducted if the final product exceeds the preset limit.

References

- Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., ... & Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, 14(1), 1983. <https://doi.org/10.1038/s41598-024-52549-8>
- Adel, A., & Alani, N. (2025). Can generative AI reliably synthesise literature? exploring hallucination issues in ChatGPT. *AI & Society*, 40(8), 6799-6812. <https://doi.org/10.1007/s00146-025-02406-7>
- Ansari, A. N., Ahmad, S., and Bhutta, S. M. (2023). Mapping the global evidence around the use of ChatGPT in higher education: a systematic scoping review. *Education and Information Technologies*, 29, 11281–11321. <https://doi.org/10.1007/s10639-023-12223-4>
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Yan, X., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *arXiv*. <https://doi.org/10.48550/arXiv.2302.04023>
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893-905. <https://doi.org/10.1080/02602938.2024.2335321>
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257-295. <https://doi.org/10.1177/23294906231176517>
- Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2024). Competencies needed by business professionals in the AI age: Character and communication lead the way. *Business and Professional Communication Quarterly*, 87(2), 223-246. <https://doi.org/10.1177/23294906231208166>
- Chappelle, C.A., & Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the test of English as a foreign language*. Routledge. <https://doi.org/10.1080/15434303.2010.519418>
- Chapelle, C. A., & Lee, H. W. (2021). Conceptions of validity. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 17–31). Routledge. <https://doi.org/10.4324/9781003220756>
- Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey. <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>
- Davenport, H., & Mittal, N. (2022). How generative AI is changing creative work. *Harvard Business Review*. <https://hbr.org/2022/11/how-generative-ai-is-changingcreative-work>
- Deng, C., Lin, S., Tong, E. K., & Ma, J. (2026). Leveraging hybrid intelligence: How teacher agency influences behavioural engagement in human-AI collaborative writing feedback practices. *International Journal of TESOL Studies*, 8 (1), 82-107. <https://doi.org/10.58304/ijts.251108>
- Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 81-83. <https://doi.org/10.54097/fcis.v2i2.4465>

- Dursun, A., & Li, Z. (2021). A systematic review of argument-based validation studies in the field of language testing (2000–2018). In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 45–70). Cambridge University Press.
- Estacio, R. D. (2025). AI integration in research writing: A quantitative study among students and teacher-researchers toward a human-AI model. *Journal of Institutional Research South East Asia*, 23(3), 373–392.
- Getchell, K., Carradini, S., Cardon, P. W., Aritz, J., Fleischmann, C., Stapp, J., & Ma, H. (2022). Artificial intelligence in business communication: The changing landscape of research and teaching. *Business and Professional Communication Quarterly*, 85(1), 7-33. <https://doi.org/10.1177/23294906221074311>
- Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection*. *arXiv*. <https://doi.org/10.48550/arXiv.2301.07597>
- Haan, K. (2023, April 4). *How businesses are using AI in 2023*. <https://www.forbes.com/advisor/business/software/ai-in-business/>
- Han, J., Yang, Y., & Liu, G. (2025). Are teachers assessing work written by students or by AI? A rapid literature review of research on detecting content generated by Generative AI. *European Journal of Education*, 60(4), e70240. <https://doi.org/10.1111/ejed.70240>
- Hwang, Y., Lee, J. H., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of new literacy skill using generative AI. *arXiv*. <https://doi.org/10.48550/arXiv.2311.05373>
- Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., et al. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports*, 13, 12187. <https://doi.org/10.1038/s41598-023-38964-3>
- Johnson, M. S. (2025). Responsible AI for measurement and learning: Principles and practices (Research Report No. RR-25-03). ETS. <https://www.ets.org/Media/Research/pdf/RR-25-03.pdf>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of educational measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Ken, H. (2026). AI feedback to L2 writers. *International Journal of TESOL Studies*, 8 (1), 146-156. <https://doi.org/10.58304/ijts.260420>
- Kim, J., Yu, S., Lee, S. S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 1-18. <https://doi.org/10.1080/15391523.2025.2456043>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>
- Kong, A. (2021). A dual case study of perceptions of strategy-based training and peer review stances among Hong Kong secondary students: An activity theory perspective. In B. L. Reynolds & M. F. Teng (Eds.), *Innovative Approaches in Teaching English Writing to Chinese Speakers* (pp. 43-67). De Gruyter Mouton. <https://doi.org/10.1515/9781501512643>
- Kong. (2024). Examining the argument-based validity of the HKDSE speaking test by investigating the domain definition, explanation, and utilization inferences. *Asian Journal of English Language Teaching*, 33(1), 53-86. <https://doi.org/10.65961/AJELT-2024-1-004>

- Kong, A., & Bui, G. (2019). Reader stances and writer responses in L2 peer review: A study of L2 writing literacy among Hong Kong secondary school students. *Asian EFL Journal*, 21(3), 140-186. <https://www.asian-efl-journal.com/main-journals/2019-main-journal/volume-23-issue-5-2019/>
- Kong, A., & Teng, M. F. (2023). The operating mechanisms of self-efficacy and peer feedback: An exploration of L2 young writers. *Applied Linguistics Review*, 14(2), 297-328. <https://doi.org/10.1515/applirev-2020-0019>
- Lee, S.S., Li, N., & Kim, J. (2024). Conceptual model for Mexican teachers' adoption of learning analytics systems: The integration of the information system success model and the Technology Acceptance Model. *Education and Information Technologies*, 29(11), 13387–13412. <https://doi.org/10.1007/s10639-023-12371-7>
- Legislative Council Secretariat. (2025, October). *Measures to support development of AI education tools in the United Kingdom and Singapore* (Essentials ISE11/2025). Research Office, Research and Information Division, Legislative Council Secretariat.
- Li, H., & Pan L. (2024). Network of discourses: resistance and negotiation within Chinese students' AI-assisted EFL writing. *International Journal of TESOL Studies*, 6 (3), 128-145. <https://doi.org/10.58304/ijts.20240309>
- Lin, H. C. K. (2025). Writing is coding: When writing becomes the starting point for AI cross-media learning and narrative. *International Journal of Online Pedagogy and Course Design (IJOPCD)*, 15(1), 1-23. <https://doi.org/10.4018/IJOPCD.385016>
- Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16). <https://dl.acm.org/doi/10.1145/3313831.3376727>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan Publishers.
- Mollick, E. (2023, September 16). *Centaurs and cyborgs on the jagged frontier*. Retrieved from <https://www.oneusefulthing.org/p/centaurs-and-cyborgs-on-the-jagged>
- Nelson, D. B., Gackiere, A. E., LeGrand, S. E., & Guberman, D. A. (2025). Collaborative intelligence: Towards practical, critical and cooperative teaching & learning with AI. *Thresholds in Education*, 48(1), 57-80.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192. <https://doi.org/10.1126/science.adh2586>
- Sawaki, Y., & Sinharay, S. (2013). Investigating the value of section scores for the TOEFL iBT® test. *ETS Research Report Series*, 2, i–113. <https://doi.org/10.1002/j.2333-8504.2013.tb02342.x>
- Singla, A., Sukharevsky, A., Hall, B., Yee, L., & Chui, M. (2025). *The state of AI in 2024-2025: Transitioning from experimentation to impact*. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- Park, M. (2021). Domain definition inference for virtual interactive Aviation English Test (VIAET) for military air traffic controllers. In C. A., Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 73-95). Cambridge University Press. <https://doi.org/10.1017/9781108669849.007>

- Ren, J., & Wang, W. (2025). Assisting research proposal writing with large language models: Evaluation and refinement. *arXiv*. <https://doi.org/10.48550/arXiv.2509.09709>
- Rice, R. (2025). Prompting purposefully: Teaching AI rhetorical literacy in grant writing. *Small Bytes Blog Post Submission | Fall 2025*. Retrieved from <https://www.depts.ttu.edu/tlpdc/ai-resources/small-bytes-Rich-Rice.pdf>
- Teng, M. F. (2024). “ChatGPT is the companion, not enemies”: EFL learners’ perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Teng, M. F. (2025). Understanding EFL student writers’ metacognitive awareness in utilizing ChatGPT. *System*, 135, 103848. <https://doi.org/10.1016/j.system.2025.103848>
- Teng, M. F. (2026). Validation and profiling of GenAI-mediated vocabulary learning strategies. *Asian Journal of English Language Teaching*, 35(3), 15-39. <https://doi.org/10.65961/AJELT-2026-3-002>
- von Garrel, J., and Mayer, J. (2023). Artificial intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications*, 10, 799. <https://doi.org/10.1057/s41599-023-02304-7>
- Wang, C. (2025). Exploring students’ generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*, 30(3), 1825-1846. <https://doi.org/10.1007/s10758-024-09744-3>
- Wen, H., & Hu, P. (2025). Architectural gaps in generative AI: Quantifying cognitive risks for safety applications. *AI*, 6(7), 138. <https://doi.org/10.3390/ai6070138>
- Williams, T. (2023, February 25). Some companies are already replacing workers with ChatGPT, despite warnings it shouldn’t be relied on for “anything important.” *Fortune*. <https://fortune.com/2023/02/25/companies-replacing-workers-chatgpt-ai/>
- Zapata, G. C., Saini, A., Tzirides, A. O. O., Cope, W., & Kalantzis, M. (2024). The role of AI feedback in university students’ learning experiences: An exploration grounded in activity theory. *Ubiquitous Learning*, 18(2), 1. <https://doi.org/10.18848/1835-9795/CGP/v18i02/1-30>
- Zhang, Y. (2025). GPT is all you need: AI, cognitive scaffolding, neuroplasticity and AI, AI-human interaction. *Frontiers in Psychology*, 16, 1549755. <https://doi.org/10.3389/fpsyg.2025.1549755>

Dr. Amy Kong is Senior Lecturer, English Language Centre Coordinator, and Associate Director of the MA-ELTA Programme in the Department of English at The Hong Kong University of Hong Kong. She has received multiple teaching grants from HSUHK, the Quality Enhancement Support Scheme (QESS), and the Tin Ka Ping Foundation.