



Research Papers in Language Teaching and Learning

Vol. 16, No. 1, March 2026, 13-30

ISSN: 1792-1244

Available online at <http://rpltl.eap.gr>

This article is issued under the [Creative Commons License Deed. Attribution 3.0 Unported \(CC BY 3.0\)](#)

Prompting as a Critical Literacy Practice in Higher Education: Student Reflections on Using ChatGPT for Critical Reading and Source Evaluation

Panagiota Samioti

This paper investigates how higher education students reflected on their use of ChatGPT to support critical reading and source evaluation within a higher education digital literacy course in Greece. Drawing on students' written reflections, forum discussions, collaborative tasks and questionnaires, the study explores how students interacted with ChatGPT as a dialogic partner for questioning, verification, and reasoning. Thematic analysis of the students' data identified four interrelated themes: (1) prompting as a metacognitive practice, (2) developing a critical stance through evaluative judgment of ChatGPT's outputs, (3) recognizing the tool's conditional usefulness, and (4) addressing language-related challenges. Findings show that students demonstrated awareness of how the formulation of prompts influenced response quality and depth, using inaccuracies and generalizations as opportunities for deeper analysis. They valued ChatGPT's immediacy and accessibility but maintained a cautious stance toward its reliability, highlighting the need for verification and human mediation. Differences in performance across languages also prompted reflection on linguistic precision and bias. Overall, students' reflections highlight that structured, reflective prompting can foster metacognitive regulation, evaluative judgment, and critical reading, positioning ChatGPT as a catalyst for inquiry and ethical awareness rather than a substitute for human reasoning.

Keywords: ChatGPT; critical reading; source evaluation; metacognition; AI literacy; higher education; language awareness

1. Introduction

The rapid emergence of generative artificial intelligence (GenAI) has reshaped the educational landscape, raising urgent questions about pedagogy, ethics, and academic practice. Among these tools, ChatGPT has gained prominence in higher education, where it is used to support feedback, scaffolding, and reflective learning. As its presence expands, understanding how

students critically engage with such technologies becomes essential for developing responsible pedagogical frameworks that enhance human judgment and interpretive skills.

This study addresses this need by examining higher education students' reflections on using ChatGPT (3.5, free version) (OpenAI, 2024/5), as it was available at the time of course implementation within a European-funded higher education project (2022-2025). Although the course centered on academic digital literacy rather than explicit language instruction, it required students to engage in critical reading and source evaluation, practices that are closely tied to language awareness and critical literacy in academic settings. In this study, higher education is treated as a key site of language education, particularly in English for Academic Purposes (EAP) and academic literacy contexts, where critical reading and evaluative judgment are central pedagogical goals. The tasks that supported critical reading and source evaluation were inherently language-based, as they involved interpreting meaning, analyzing discourse, and considering how language choices shape credibility and persuasion in academic communication. Positioning ChatGPT within these language-focused activities allowed students to experience AI not merely as a reading tool but as a medium for reflection, interpretation, and evaluative judgment demonstrating how AI-mediated interaction can help learners question, verify, and interpret meaning across genres and languages in language education contexts.

Specifically, the research examines how students employed ChatGPT to interrogate academic and non-academic texts, identify argumentative and linguistic techniques, and evaluate issues of bias, authorship, and reliability. ChatGPT was not presented as a provider of answers but as a critical interlocutor, that is a tool for prompting inquiry, comparison, and reflection. This framing conceptualizes prompting as both a metacognitive and critical practice through which students cultivate awareness of how language constructs meaning and persuasion. In this study, prompting is understood not as a technical skill but as a critical literacy practice embedded in language-mediated academic activity.

In this study, critical reading and source evaluation are treated as complementary components of critical literacy. While critical reading involves analyzing stance, argumentation, and rhetorical strategy, entailing the ability to identify claims, evaluate reasoning, and interpret perspectives (Flemming, 2012; Barnett & Bedau, 2011, both cited in Sultan *et al.*, 2017) source evaluation focuses on assessing the credibility, bias, and evidential quality of information. Together, they nurture critical linguistic awareness as the ability to interpret and use language reflectively and ethically, which bridges academic literacy and AI literacy.

Drawing on written reflections, forum discussions, collaborative tasks and questionnaires, the analysis identifies four interrelated themes: (1) prompting as a metacognitive practice, (2) critical engagement with AI-generated content, (3) conditional usefulness of GenAI, and (4) language issues in AI outputs. The study therefore aims to explore how structured, reflective use of GenAI, and more specifically ChatGPT's use, can prompt a critical stance enacted through evaluative judgment and verification and support students' development of evaluative judgment, metacognitive regulation, and ethical awareness in higher education. In doing so, it contributes to ongoing debates in applied linguistics and language education about how GenAI can be pedagogically embedded to promote critical reading, source evaluation, and human-centered reasoning in an AI-mediated academic environment.

2. Literature Review

2.1 AI and Academic Literacy

Studies converge on the view that GenAI, such as ChatGPT, can support academic literacy, but its effectiveness depends on the conditions of use. So far, much research has concentrated on writing-related affordances. For instance, in a study, students were enthusiastic in using ChatGPT for essay writing (Šedlbauer, 2024) and, in another, students have appreciated AI support for writing fluency (Yuan et al., 2024) and for content and organization (Mahapatra, 2024; Yuan et al., 2024). Mendez & Tang (2025) report that Swedish students valued its contribution to lower-level outcomes such as summarization and editing. Similarly, positive attitudes have been reported regarding grammar assistance (Johnston et al., 2024; Mahapatra, 2024). However, improvements appear most evident when learners already possess strong critical and ethical foundations, as Baldrich and Domínguez-Oller (2024) note.

Although writing-related affordances dominate literature, research has also explored broader dimensions of academic literacy support, including brainstorming, scaffolding, lecture comprehension, and the development of learner autonomy. Thus, research shows that students expressed appreciation for brainstorming support (Chan & Hu, 2023) and for scaffolding content and knowledge, especially in EFL contexts (Yuan et al., 2024). Undergraduate and postgraduate students have also acknowledged AI's benefits in enhancing lecture comprehension (Kostas et al., 2025), and positive attitudes were reported toward personalized support and language assistance (Chan & Hu, 2023). Also, Mahapatra (2024) underscores ChatGPT's value as a dialogic feedback tool, particularly in contexts where individualized feedback is limited.

Yet, despite these reported benefits, questions of learner autonomy and deeper academic processes remain central. In some studies, learners found the tool less effective for higher-order learning requiring deeper understanding, while other studies highlight risks to the development of critical thinking, academic integrity, and independent problem-solving skills, while warning that overreliance may encourage superficial engagement and academic dishonesty (Mendez & Tang, 2025; Yuan et al., 2024; Kostas et al., 2025; Pitts et al., 2025; Vieriu & Petrea, 2025; Yuan et al., 2024).

Alongside the previous questions, concerns about the quality and reliability of GenAI responses also persist. These appear in two main forms: some students judge ChatGPT's outputs as overly generic or inaccurate (Šedlbauer, 2024; Chan & Hu, 2023), while others point to cross-lingual inconsistencies, with factual queries answered correctly in some languages but less precisely in others (Xing et al., 2024). These disparities extend to accuracy and timeliness, as updates often appear first in high-resource languages like English. Xing et al. (2024) attribute these differences to uneven training data and limited cross-lingual transfer, noting that stronger translation capabilities correlate with more reliable performance. Consequently, such disparities highlight further pedagogical challenges. On the one hand, working in underrepresented languages may require learners to invest more effort in crafting precise prompts or verifying outputs, potentially enhancing metacognitive and linguistic awareness. On the other hand, inequities risk disadvantaging non-English users. Such inconsistencies complicate students' abilities to summarize texts accurately, evaluate sources, and transfer academic reading or writing strategies across languages.

These challenges reinforce the importance of pedagogical mediation and responsible integration of GenAI technologies to scaffold academic literacy. Research stresses that GenAI

should remain embedded in pedagogical frameworks that preserve critical engagement and human interaction. For instance, in English for Academic Purposes settings, Du & Alm (2024) observe that ChatGPT offers flexible opportunities for practice and supports competence, yet they emphasize that teacher-student interaction remains essential for meaningful literacy development. Extending this pedagogical perspective, Anson (2024) argues that large language models (LLMs) can stimulate reflection by generating counterpoints or scaffolds for argument development when positioned as complementary tools. Similarly, another study shows that explicit training in AI use, combined with self- and peer-assessment, helps ensure that ChatGPT supports reflection and deeper engagement with writing rather than serving as a shortcut (Mahapatra, 2024). Finally, at a broader level, recent research highlights the need for institutional policies that promote responsible academic writing practices and uphold pedagogical integrity (Johnston et al., 2024).

2.2 Critical Reading and Source Evaluation with AI

Although extensive research has addressed writing, autonomy, and reliability, fewer studies have examined how AI shapes source evaluation and critical reading, an area to which the present study contributes. In this context, the term “critical reading” is used broadly to encompass students’ evaluative and analytical engagement with AI-generated content, including their reasoning, verification, and interpretation of information credibility and relevance.

Within applied linguistics, critical reading is understood as a language-mediated practice through which readers examine how meaning, authority, and interpretation are constructed in texts. Drawing on critical pedagogy, Wallace (2003) conceptualizes critical reading as close attention to linguistic and grammatical choices as resources for reflecting on ideological positioning, credibility, and social context. Extending this view to AI-mediated contexts foregrounds the role of language in evaluating how claims, perspectives, and sources are framed. From a second language pedagogy perspective, critical literacy positions texts as socially situated and open to interrogation, emphasizing learner agency (Luke & Dooley, 2011). Similarly, in EAP, reading is an evaluative practice assessing stance, evidence, and disciplinary credibility (Hyland, 2004).

When viewed through these language-education lens, research on generative AI tools such as ChatGPT indicates that they can scaffold analytical reasoning and source analysis but also risk fostering overreliance and superficial engagement. This duality is evident in studies that frame AI as a collaborative teammate in problem-solving contexts, where students valued its feedback but often overestimated its reliability and cognitive capacity (Marrone et al., 2025). Such misconceptions parallel those observed in critical reading tasks, where uncritical trust in AI-generated information can hinder independent verification and evaluative reasoning.

Nevertheless, when embedded within structured pedagogical strategies, ChatGPT can promote deeper critical engagement. Research shows that EFL learners critically evaluated its responses by posing follow-up questions and verifying information, strengthening understanding and critical thinking (Liang & Wu, 2024). Students also gained confidence in probing questions, analysis, and conceptual understanding, with diverse perspectives challenging assumptions and highlighting the need for refined prompts (Guo & Lee, 2023). Similarly, constructing scientific argument maps around ChatGPT outputs enabled learners to anticipate counterarguments and develop rebuttals, enhancing argumentation skills (Archila et al., 2025). In language education, reflective journals indicate that ChatGPT supported

spatially aware, multi-perspective critical thinking while underscoring the importance of balancing technological support with human reasoning (Liang & Wu, 2024). Finally, students frequently described ChatGPT as a supportive “study partner” offering guidance and feedback (Chan & Hu, 2023; Šedlbauer, 2024).

However, these affordances are constrained by persistent challenges that complicate the picture. For example, research warns that excessive reliance on ChatGPT can discourage students from verifying accuracy or sources, thereby weakening their ability to evaluate authenticity and critical thinking in academic writing, and may even lead to ‘cognitive dependence’ when outputs are accepted uncritically with little engagement (Yuan et al., 2024; Sari et al., 2025; Suriano et al., 2025). In addition, students may encounter inaccuracies, bias, or fabricated content in AI outputs (Archila et al., 2025; Harrer, 2023). Survey-based studies reinforce these concerns, documenting risks to integrity and critical thinking (Kostas et al., 2025). Finally, many students perceived GenAI outputs as generic (Šedlbauer, 2024), while research warns that efficiency gains with the use of AI technologies may come at the expense of critical depth and authorship clarity (Vieriu & Petrea, 2025; Mendez & Tang, 2025).

These findings reveal a recurring paradox; ChatGPT can foster critical reading and source evaluation skills, when embedded in reflective, teacher-mediated tasks, but risks reinforcing surface-level learning and undermining autonomy when used uncritically. Therefore, effective integration requires explicit attention to AI literacy, institutional guidance, and pedagogical design that encourage students to question, compare, and engage in critical evaluation of outputs rather than passively accepting them (Walter, 2024).

2.3 Prompt Crafting and Metacognitive Awareness

Metacognitive regulation, that is planning, monitoring, and evaluating one’s own thinking processes, has long been recognized as essential for effective learning (Flavell, 1987; Schraw & Dennison, 1994, both as cited in Teng, 2024). However, students who depend on AI answers without critical reflection risk falling into “metacognitive laziness” (Sari et al., 2025). This underscores the importance of positioning AI not as an answer-giver but as a dialogic partner, an approach that becomes especially salient in the context of GenAI, where these regulatory processes are most visible through prompt crafting.

Recent studies suggest that effective engagement with ChatGPT depends on learners’ reflection on how prompt phrasing influences the relevance and accuracy of responses. Reflective engagement is strengthened when students recognize AI limitations, such as difficulty capturing irony or producing plausible but inaccurate outputs, which often trigger fact-checking, source triangulation, and cautious use (Darwin et al., 2024; Essien et al., 2024; Emran et al., 2024, as cited in Raitskaya & Tikhonova, 2025). Iterative prompting and verification further reinforce monitoring behaviors and metacognitive regulation (Raitskaya & Tikhonova, 2025). Metacognition is therefore central to learners’ use of AI feedback (Teng, 2024). Finally, empirical evidence shows that sustained experimentation with prompts promotes higher-level cognitive engagement and deeper critical thinking (Kavadella et al., 2024), especially when guided, as structured mediation can transform superficial AI outputs into opportunities for reflection, elaboration, and higher-order thinking (Borge et al., 2024).

Building on this, research highlights the pedagogical value of structured support in prompting. Guo and Lee (2023) emphasize the importance of crafting clear and specific prompts to obtain more relevant and higher-quality responses. In their study, they implemented a three-stage

ChatGPT-based activity in chemistry courses, that is orientation, essay creation, and output validation, which allowed students to decompose tasks, refine prompts, and verify information. The results showed that detailed instruction was not only crucial for success but also helped students gain confidence in analyzing, evaluating, and drawing logical conclusions. Raitskaya & Tikhonova (2025) also reported that structured interventions around prompt use can cultivate students' critical awareness by requiring them to verify AI-generated responses against other sources. This form of guided practice complements approaches described in their review, such as Lee et al.'s (2024, as cited in Raitskaya & Tikhonova, 2025) guidance-based tool, which used indirect prompts to encourage learners to articulate their reasoning, and Hwang et al.'s (2025, as cited in Raitskaya & Tikhonova, 2025) prompt-based learning model, which strengthened students' ability to generate questions and reflect on their learning processes. Together, these studies underscore that structured interaction with AI, when mediated through carefully designed instruction, enhances both the quality and depth of students' reflective thinking.

Overall, research shows that prompt crafting is a metacognitive practice fostering reflection, regulation, and rhetorical awareness; with structured support and epistemic vigilance it enables deeper learning and critical engagement, but without scaffolding it risks metacognitive laziness, making explicit instruction in prompt formulation and interrogation essential.

3. Methodology

3.1 Research Design

The present study adopted a mixed-methods research design to examine student reflections on the use of GenAI in higher education. The data were produced during the implementation of a hybrid (both in situ and online) six-session academic digital literacy course delivered within a European-funded higher education project (spring semesters 2024 and 2025) focused on critical reading and source evaluation. It is important to emphasize that the course itself was not designed as a research intervention. For the purposes of the present research, the data generated during the course activities were subsequently analyzed to explore how students engaged with ChatGPT.

Specifically, the course aimed to strengthen students' ability to identify the features of credible versus misleading information, distinguish between academic and non-academic or false sources, analyze persuasive and rhetorical language, and reflect on the responsible use of AI tools. Students worked with a variety of materials, including academic journal articles, news media, and social media content, and engaged with digital tools such as the Truly Media Platform (Athens Technology Center & Deutsche Welle., n.d.) and ChatGPT 3.5 (free version) (OpenAI, 2024, 2025). Forty-six undergraduate and postgraduate students from multiple University departments participated. Instruction was in Greek; materials were in Greek and English; responses were in Greek. All were Greek L1 speakers, enabling cross-linguistic reflection on ChatGPT. Despite disciplinary diversity, all developed critical digital literacy through structured engagement with academic and non-academic texts.

3.2 Data Collection and classroom tasks

Four types of data were analyzed:

1. Written reflections: Individual or group reflections on students' experiences with ChatGPT during class activities/tasks.
2. Forum written discussions: Online debates about the reliability and ethical implications of AI-generated content.
3. Collaborative written tasks: Group products from source analysis activities and public-facing texts created to counter misinformation.
4. Quantitative data from final course evaluation questionnaires.

Four datasets were analyzed, each derived from the following activities:

- Dataset 1: Group reflections on an activity where students asked ChatGPT which misinformation-related terms (e.g., propaganda, clickbait, manipulation, deepfakes) were most important for students, scientists, and citizens. They compared its responses with their own reasoning and shared their reflections in the forum.
- Dataset 2: Analyses of persuasive strategies in various text genres (e.g., journalistic, multimodal). Students input these texts into ChatGPT to identify rhetorical techniques such as appeals to authority, logic, and emotion.
- Dataset 3: Tasks where students examined contested or fabricated sources across genres. One case involved the academic "chocolate diet" hoax article (Bohannon, 2015), where ChatGPT assessed claim validity and evidence. Other groups explored its evaluations of bias, exaggeration, and fallacies in texts on technology addiction, social media disinformation, and climate news.
- Dataset 4: Evaluations of argumentative writing and the role of language as a persuasive device. ChatGPT was asked to assess whether texts employed emotional vocabulary, hyperbole, or bias, and students compared its outputs with their own judgments.

3.3 Data Analysis

The data were analyzed using thematic analysis (Braun & Clarke, 2006), following a six-phase process of familiarization, coding, theme generation, reviewing, defining, and reporting. Credibility was enhanced through both data-type and methodological triangulation, combining qualitative and quantitative sources to ensure a comprehensive interpretation of findings. Coding was refined iteratively, and thick descriptions of student voices were maintained to preserve authenticity.

An inductive-deductive approach was adopted to capture both emergent and theory-informed patterns in the data. Inductive coding allowed themes to emerge directly from students' reflections and discussions, representing their authentic perspectives. Deductive coding, in turn, was guided by key constructs identified in the literature, such as critical reading and source evaluation, metacognitive awareness and prompting, curiosity and exploratory engagement, autonomy and responsible use, and ethical reflection.

This dual process resulted in four themes: (1) formulating prompts as a reflective and metacognitive practice, (2) developing a critical stance through evaluative judgment of ChatGPT's outputs, (3) recognizing the conditional value of ChatGPT as a learning tool, and (4) addressing language-related challenges and cross-lingual variability in AI performance.

All coding and theme development were conducted by the author, who also taught the course. Although intercoder reliability was not possible, rigor was ensured through prolonged engagement, iterative coding and refinement, systematic data triangulation, and reflexive memoing to address the author's dual instructional and research role.

3.4 Limitations

Several limitations must be acknowledged. First, the academic digital literacy course was a pedagogical initiative within a European-funded project rather than a research intervention. While this enhances ecological validity by capturing authentic classroom practices, it limited design control, as there were no pre- and post-measures, comparison group, or experimental manipulation; therefore, claims about change over time should be treated cautiously. Second, findings derive from a relatively small, context-specific sample from a single Greek higher education course. Although triangulation across reflections, forum discussions, collaborative tasks, and questionnaires strengthens credibility, results are not generalizable beyond this setting. Third, despite data from two cohorts (2024–2025), the study was not longitudinal, and references to increased awareness reflect recurring patterns rather than measured development. Fourth, thematic analysis involves interpretive decisions; transparency and extensive quotations mitigate alternative interpretations. Finally, the author’s dual instructor–researcher role may introduce bias, addressed through pattern-focused analysis and verbatim evidence.

3.5 Ethics

The research was conducted in line with institutional and Erasmus+ ethical standards. Students were informed that their course output might be used for research purposes, and informed consent was obtained from all participants. Participation was voluntary, and all data were anonymized before analysis to protect student identity. No identifying details are included in the findings. Since the data were originally generated in a pedagogical setting and later repurposed for research, special care was taken to ensure that consent, confidentiality, and respect for participants’ voices were maintained throughout the process. Finally, students were informed about data privacy and terms-of-service considerations when using ChatGPT.

4. Findings

The analysis of written reflections, forum discussions, and collaborative tasks revealed four interrelated themes: (1) formulating prompts, (2) development of critical stance, (3) usefulness with limits, and (4) language issues. Each theme is illustrated through representative quotations that capture key patterns in the data, while only selected student excerpts are presented.

4.1 Prompting as Metacognitive Practice

A strong theme concerns the prompt design process. Students consistently described ChatGPT as useful yet dependent on their own mediation. In a group discussion activity, one group reflected that “with our guidance we can get valuable answers from the ChatGPT concerning arguments” pointing to the need for active involvement. Also, one group highlighted that “the answer changes depending on the role we assign to ChatGPT (e.g., teacher or journalist)”, showing how framing influenced responses.

Many students explicitly recognized that the way they phrased their questions shaped the quality of ChatGPT’s answers. As one group put it: “The way the question is formulated plays an important role for a clearer and more targeted answer”. Another added that “in some cases,

the answers were generalized, while in others more specific depending on the way the question was expressed". Several groups remarked that ChatGPT gave only "general" answers unless their questions were very specific, with some students noting that "the more precise and specific questions we pose, the more personalized and topic-relevant answers it will give", confirming the link between prompt quality and output.

4.2 From Iterative Prompting to Evaluative Judgment: Developing a Critical Stance

A second major theme is the development of critical stance. Students' awareness of the relationship between prompt quality and response depth was reflected in a more deliberate and reflective stance toward their questioning practices. They designed systematic processes to elicit more meaningful and accurate responses from ChatGPT. For instance, when analyzing the "chocolate diet" hoax article (Bohannon, 2015), one student explained:

First, I asked ChatGPT if it knew the techniques of persuasion, and it responded with a specific list. Then I provided the text and asked it to analyze the persuasive techniques. Others described similar step-by-step strategies, such as dividing long texts into smaller parts so that the tool could process them and clarify its understanding: *I had to open a new chat and put in half the article to analyze and then the other half or posing the task's questions one by one "in the form of questions as they were in the Word document" to obtain more balanced and nuanced answers.* Some even highlighted the importance of metacognitive control over the process, noting that "the more precise and specific questions we pose to ChatGPT, the more personalized and topic-relevant answers it will give us". Through such iterative experimentation, students transformed prompt formulation into a deliberate, reflective, and strategic practice rather than a mechanical query-response routine.

Students constantly reflected on the answers that ChatGPT gave. For example, in analyzing the "chocolate diet" hoax article (Bohannon, 2015), they reported that ChatGPT pointed out "logical fallacies... missing details about methodology, no alternative interpretations, and lack of independent confirmation". Students noted that ChatGPT correctly revealed it as a deliberate hoax, which led them to reflect on "how easily scientific articles can be published without strict peer review". As one student concluded, such analyses underscored "the importance of critical thinking and proper evaluation of scientific sources".

Even though students valued ChatGPT's outputs, they positioned themselves as active evaluators rather than passive recipients. Their reflections show that when ChatGPT identified weaknesses in arguments, they treated these not as final answers but as opportunities to sharpen their own critical thinking. For example, students contrasted ChatGPT's answers with their own perspectives, thereby exercising judgment. One student remarked that while the chat claimed an article presented "many logical arguments" their own view was that "the article mentions opinions and research but does not really argue strongly". Another student observed that ChatGPT tended to downplay bias, treating sources as more balanced than they appeared. Finally, one said: "ChatGPT answered that there are a lot of logical arguments, whereas my personal view is that the article mentions opinions and studies but does not argue that much".

4.3 Conditional Usefulness

Regarding the third theme, a group reported that “the answers ChatGPT gave were quite informative and specific” while another highlighted that, even when responses were not fully adequate, “it covered the issue, it just cannot show 100% reliability». Others similarly described the interaction as “quick and clever” but still requiring their own critical thinking.

At the same time, students expressed doubts about ChatGPT’s reliability in academic contexts. A recurring issue was inconsistency: “different users received different answers to the exact same question” with responses shifting between generalized statements and more detailed elaborations depending on how prompts were phrased. Also, some students noted misleading or fabricated content. In one case, a group reported that ChatGPT “invented a source that does not exist”, while another student was more explicit: *I believe it is very easy to be misled regarding the reliability of a source. Appearances are certainly deceiving.*

As one student concluded: *I think it is a good way at a first stage... but in no case should we trust it, and we must always re-examine the validity of what it states.* Another emphasized that *“regarding reliability, we always need to check sources and not rely only on ChatGPT”.* Finally, another added: *“The answer was quite satisfactory, but it requires further investigation”* signaling recognition that ChatGPT could be a useful starting point but shouldn’t be considered a unique or definitive source.

Taken together, these qualitative accounts already frame ChatGPT as conditionally useful: valued for speed and initial orientation, but consistently positioned by students as requiring verification, critical judgment, and follow-up inquiry.

This qualitatively articulated cautious stance was also reflected in the final course evaluation questionnaire results (Figure 1), based on the item that asked students: “Is ChatGPT a reliable tool to be used in academic essays?”. Responses were almost evenly split: 16 students said “Yes” 14 said “No” and another 16 qualified their answer by saying “It depends on the type of academic essay”. Such divided opinions underline the perception of ChatGPT as conditionally useful rather than universally reliable.

The questionnaire item (Figure 2) in the same questionnaire that asked students: “To what extent do you trust the information provided by ChatGPT?” reinforces this picture. The vast majority of students (43 out of 46) selected “So and so” indicating a moderate level of trust that depends on context. None reported trusting it “Not at all” while only three students said they “Absolutely” trusted the tool. This distribution confirms that students were willing to engage with ChatGPT but remained skeptical about relying on it without verification.

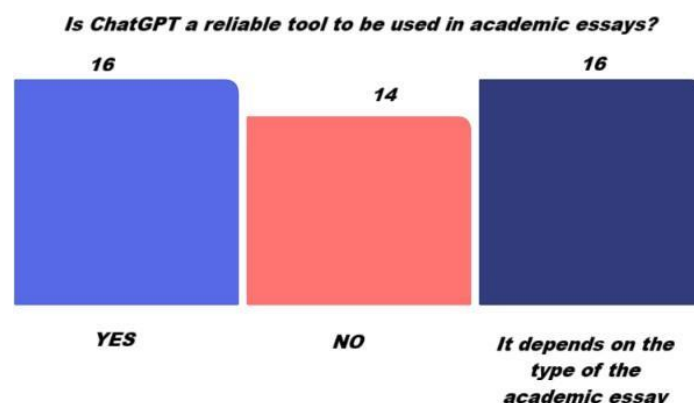


Figure 1: Perceptions of ChatGPT reliability

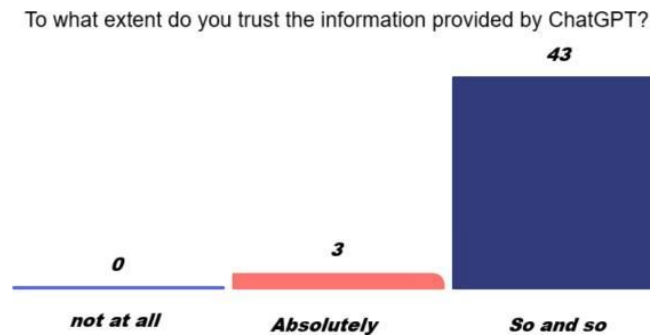


Figure 2: Perceptions of ChatGPT trustworthiness

Taken together, these accounts show that students appreciated ChatGPT's speed, clarity, and perceived usefulness, but they also recognized its limitations in terms of reliability and credibility. Such reflections underline a cautious stance toward integrating AI-generated content into academic tasks, where reliability and critical evaluation remain essential.

4.4 Language-Related Challenges

Several students observed that ChatGPT's performance varied across languages. They noted that responses in English tended to be more complete and more precise, which they attributed to the larger training data available in that language. By contrast, in Greek the tool often struggled, requiring users to phrase questions more carefully in order to obtain meaningful answers.

A group explicitly noted problems with Greek since ChatGPT mistranslated "manipulation" as "transformation" and confused "citizens" with "politicians". Another group observed: *In general, it seems that ChatGPT has difficulty with questions in Greek and it needs the questions to be more specific, or for us to ask more questions.*

Finally, some students remarked that, although ChatGPT identified exaggerated or figurative language, "in the linguistic domain it falls short and in no case can we take it into account without critical thinking". Students also reflected on the tone and style of ChatGPT's outputs. Some characterized the responses as "diplomatic, leaving room for further investigation", while others found them overly "robotic", often presented in bullet points or summary form. They noted that "searching in English is more complete and clearer... also less stereotyped and robotic". These observations highlight how language choice and stylistic features influenced their perceptions of ChatGPT's usefulness.

5. Discussion

This study's findings reveal how higher education students used ChatGPT to evaluate sources and cultivate critical reading, highlighting the pedagogical potential of AI-mediated prompting. Viewed through a language-education lens, these practices reflect core concerns of second language reading and critical literacy, including evaluative judgment, metalinguistic awareness, and attention to how language constructs credibility and authority.

Four interrelated topics emerged; prompting as a metacognitive practice, critical engagement, conditional usefulness, and language variability, each revealing how engagement with

ChatGPT can cultivate critical literacy as a co-thinking process rather than a substitute for independent reasoning.

5.1 Prompting as a Metacognitive Practice

Regarding the first topic, students recognized the central role of prompt design and iteration in shaping the specificity and depth of ChatGPT's responses following a pattern which resonates with classic conceptualizations of metacognition; the active planning, monitoring, and evaluation of one's cognitive processes (Flavell, 1987; Schraw & Dennison, 1994, as cited in Teng, 2024). Furthermore, they emphasized that the clarity and framing of their questions determined the quality of the output, reflecting metacognitive awareness of how linguistic choices and question framing shape meaning. This finding aligns with Teng's (2024) argument that metacognitive regulation is integral to AI-supported learning and with Kavadella et al.'s (2024) observation that iterative prompting enhances reflection and higher-order thinking. It also supports Raitskaya & Tikhonova's (2025) view that guided interventions around prompt use promote epistemic vigilance, encouraging learners to monitor and evaluate their interactions critically. Similarly, Lee et al. (2024, as cited in Raitskaya & Tikhonova, 2025) and Hwang et al. (2025, as cited in Raitskaya & Tikhonova, 2025) emphasize that indirect and structured prompts can scaffold students' reasoning processes and deepen reflective engagement. Taken together, these findings demonstrate that deliberate and well-structured prompting nurtures metacognitive control and fosters sustained reflection during AI interaction.

5.2 Critical Engagement and Evaluative Judgment

While the first aspect of metacognition focused on deliberate prompt design, the second was evident when students engaged with ChatGPT's limitations which stimulated additional reflection and evaluation. Similarly to what Darwin et al. (2024), Essien et al. (2024), and Emran et al. (2024) note, the learners in the present study became more reflective when they encountered AI's limitations, turning these shortcomings into opportunities for verification and refinement. More specifically, students reflected on the limitations of ChatGPT's shallow or bulleted outputs and refined them, confirming Borge et al.'s (2024) assertion that ChatGPT can be a simulation tool for higher-order thinking. Therefore, students' active experimentation and reflection counter the risks of "metacognitive laziness" noted by Sari et al. (2025), showing that structured AI use can strengthen both critical reading and metacognitive skills.

Having examined prompting as metacognition, the next theme addresses how critical engagement unfolded through evaluative judgment. Students identified inaccuracies, vague reasoning, and fabricated references in ChatGPT's responses, treating these not as failures but as prompts for deeper analysis. For example, in evaluating the "chocolate diet" hoax article (Bohannon, 2015), they validated ChatGPT's detection of flawed methodology while independently assessing the credibility of the evidence. This engagement supports Liang and Wu's (2024) and Guo & Lee (2023) findings that students can use ChatGPT to stimulate probing questions and strengthen analytical reasoning. It also echoes Archila et al. (2025), whose work demonstrates how engagement with ChatGPT can enhance reasoning and evidence evaluation, showing that these processes can be equally central to critical reading and source assessment. Extending their critical engagement further, students also recognized broader issues of bias and reliability. Their identification of inaccuracy and occasional fabrication in ChatGPT's outputs reflects concerns noted by Harrer (2023), who warns that AI-generated content often reproduces distortions that must be critically interrogated. Students'

recognition of ChatGPT's diplomatic phrasing and tendency to generalize illustrate further their capacity for reflective critique, exemplifying the kind of evaluative stance that Yuan et al. (2024) and Suriano *et al.* (2025) identify as essential to avoid cognitive dependence. These insights are also consistent with Pitts et al. (2025), who caution that uncritical acceptance of AI output may lead to superficial engagement, underscoring the importance of reflective scaffolding. Collectively, these findings confirm Walter's (2024) assertion that AI can foster critical literacy when learners engage in reflective dialogue rather than accept outputs uncritically. Overall, these examples suggest that students used ChatGPT's limitations, that is its occasional superficiality, tendency toward diplomatic answers, and generalizations, as opportunities to interrogate responses, compare them with other evidence, and reflect on standards of reliability and argumentation. In this sense, the tool became a catalyst for sharpening students' evaluative judgment and critical reading skills.

These practices are central to EAP (Hyland, 2004) and second language reading pedagogy, aligning with reading-to-write research that treats locating, evaluating, and using information as core academic practices (Council of Writing Program Administrators, 2014; Grabe & Stoller, 2011; Horning, 2017, as cited in Kocatepe, 2021). From this perspective, students' engagement with ChatGPT reflects established academic reading expectations, where evaluative judgment and source scrutiny are integral to meaning-making and knowledge construction.

5.3 Conditional Usefulness of ChatGPT

The third theme, conditional usefulness, revealed students' nuanced understanding of ChatGPT's strengths and limitations. While they valued its immediacy, clarity, and accessibility, they also emphasized the need for verification and human mediation.

Questionnaire data showed divided trust, with many describing ChatGPT as a useful starting point but not a definitive source. This cautious optimism reflects earlier research reporting both enthusiasm for idea generation and concerns about reliability (Šedlbauer, 2024; Chan & Hu, 2023; Mahapatra, 2024). Following Anson (2024) and Mahapatra (2024), students positioned ChatGPT as a supplementary aid that supports reflection rather than replaces analytical reasoning. At the same time, as Vieriu and Petrea (2025) note, the efficiency gains offered by AI can reduce critical depth and blur authorship boundaries, a limitation that students in this study appeared to address through verification. More specifically, by emphasizing mediation and verification, students demonstrated the form of informed critical judgement envisioned in the course objectives.

5.4 Language Variability and Cross-Linguistic Awareness

The final theme concerned cross-linguistic variability and inequity in ChatGPT's performance. Students observed that English outputs were more complete, precise, and stylistically natural than those produced in Greek, often noting mistranslations, semantic ambiguities, or culturally awkward phrasing in the latter. This observation corroborates findings by Xing et al. (2024) and Kostas et al. (2025), who attribute such disparities to uneven training coverage that privileges high-resource languages like English and limits representational balance across linguistic contexts. Yet rather than viewing these limitations merely as technical flaws, students used them as occasions for critical linguistic reflection and inquiry, adjusting prompts, comparing answers across languages, and refining their phrasing to elicit clearer responses.

These adaptive strategies reflect emerging multilingual AI literacy and heightened metalinguistic awareness, as learners became more conscious of how language mediates meaning, accuracy, and bias in AI interaction. By foregrounding this active negotiation of meaning across languages, the findings extend current understandings of critical literacy in AI-mediated environments and reveal how linguistic variability—though a constraint—can also function as a pedagogical resource. Students’ reflections illustrate that such variability encouraged them to interrogate issues of representation, linguistic equity, and the sociocultural dimensions of AI communication, thereby linking language awareness with ethical and critical engagement. These findings align with broader discussions in language education that advocate for ethical, reflective, and pedagogically mediated use of generative tools (Crompton et al., 2024; Du & Alm, 2024; Walter, 2024; Anson, 2024), highlighting that structured classroom tasks and guided mediation play a decisive role in transforming linguistic limitations into opportunities for inquiry and metacognitive growth. When supported by such scaffolding, cross-linguistic challenges become productive sites for reflection on language, bias, and interpretation, reinforcing the view that AI integration in higher education must be designed not only for efficiency but also for linguistic diversity, critical awareness, and ethical learning.

5.5 Implications for practice in language education

ChatGPT as an answer-producing tool, the study highlights the pedagogical value of framing prompting as a language-mediated critical literacy practice embedded in academic reading and evaluation. Prompting is most effective when conceptualized as an iterative process, guiding students from broad exploratory questions to more focused prompts targeting argumentation, evidence, sources, or rhetorical strategies (see Appendix A). Such sequencing foregrounds how prompt formulation shapes analytical depth and supports metacognitive awareness.

AI-mediated prompting is also most effective when integrated into structured critical reading tasks. Comparing ChatGPT’s outputs with students’ own analyses encourages identification of convergence, omission, and disagreement, positioning ChatGPT as a dialogic resource subject to evaluation and verification. In addition, explicit reflection prompts play a key role in supporting metacognitive regulation by helping students articulate how rephrasing prompts alters responses, what assumptions AI makes, and which claims require external verification.

Finally, cross-linguistic and genre-based prompting fosters critical language awareness by surfacing issues of precision, bias, and contextual appropriateness. Overall, effective integration of ChatGPT in language education depends on conceptualizing prompting as a critical literacy practice that is deliberately designed, scaffolded, and reflected upon, aligning with broader goals of academic literacy and responsible AI use.

6. Conclusion

The data portray students as cautious yet reflective users who balanced appreciation of ChatGPT’s affordances with critical awareness of its epistemic and linguistic limitations. These reflective practices were fostered through structured tasks and guided activities that positioned ChatGPT as a dialogic partner rather than an authority. Within this pedagogically mediated framework, students often acted as co-mediators, rephrasing prompts, verifying information, and evaluating the reliability of outputs. As several noted, “with our guidance we

can get valuable answers”, acknowledging that effective engagement depended on human judgment and instructional framing. From a language-education perspective, these practices align with pedagogical goals in second language reading and English for Academic Purposes (EAP), where critical evaluation of sources, stance, and evidence constitutes a core outcome of academic literacy rather than a peripheral skill. This pattern supports Du and Alm’s (2024) argument that meaningful AI integration relies on pedagogical mediation that sustains purposeful human-AI dialogue and teacher-student interaction. Although direct teacher input was not always visible, it was embedded in the task design that guided students toward reflective inquiry. Such design foregrounded critical reading, evaluative judgment, and learners’ agency in meaning-making across languages, illustrating how AI-mediated tasks can be aligned with language-pedagogical priorities rather than efficiency-driven uses of technology. Related research (Johnston et al., 2024) likewise highlights students’ need for clear institutional and instructional guidelines, underscoring that pedagogical mediation, whether through teachers or structured learning design, is key to responsible and critical AI use in higher education.

This study shows how generative AI can be integrated in higher education to support critical reading and source evaluation. Students’ reflective engagement with ChatGPT fostered metacognitive regulation, evaluative judgment, and critical awareness, positioning higher education, particularly EAP and academic literacy contexts, as key sites of language education and demonstrating GenAI’s potential to extend, not replace, human analytical thinking.

For educators and institutions, the implications are clear: effective AI integration requires ethical guidance, attention to linguistic and cognitive diversity, and classroom structures that promote inquiry and verification. For language educators in particular, this entails designing AI-mediated tasks that prioritize critical reading, evaluative judgment, and learners’ agency in meaning-making rather than efficiency or automation. Within such frameworks, ChatGPT can act as a catalyst for critical literacy, helping students navigate questions of authorship, credibility, and persuasion in AI-mediated academic communication.

Taken together, the findings demonstrate that structured prompting, reflective evaluation, and cross-linguistic awareness enable students to engage with ChatGPT as a partner in reasoning and reflection. By framing prompting as a reflective, language-mediated practice, the study shows how learners regulate reasoning, interrogate meaning, and evaluate epistemic claims through pedagogical mediation. It also highlights persistent linguistic inequities in AI performance, stressing the need for GenAI literacy that addresses cross-lingual variability and contributes to applied linguistics debates on ethical AI integration in EAP and second language higher education.

Appendix A

Examples of student prompts used with ChatGPT

Task / focus	Initial prompt	Refined / follow-up prompt
Identifying the main claim and evaluating argumentation	Ποιος είναι ο κύριος ισχυρισμός του συγγραφέα; (What is the author’s main claim?)	Υπάρχουν λογικά σφάλματα στο επιχείρημα του συγγραφέα; Αν ναι, ποια και πώς επηρεάζουν την αξιοπιστία του άρθρου; (Are there logical fallacies in the author’s argument? If so, which ones, and how do they affect the article’s credibility?)

Task / focus	Initial prompt	Refined / follow-up prompt
Analysing persuasive and emotionally loaded language	Το κείμενο χρησιμοποιεί λέξεις υπερβολής; (Does the text use exaggerated language?)	Μπορείς να εντοπίσεις συγκεκριμένα παραδείγματα υπερβολής ή συναισθηματικά φορτισμένου λεξιλογίου στο άρθρο και να εξηγήσεις τη λειτουργία τους; (Can you identify specific examples of exaggeration or emotionally loaded language in the article and explain their function?)
Evaluating source credibility and evidence	Είναι αξιόπιστη αυτή η μελέτη; (Is this study reliable?)	Αξιολόγησε την αξιοπιστία της μελέτης λαμβάνοντας υπόψη το περιοδικό, τους συγγραφείς και τη μεθοδολογία. (Evaluate the credibility of the study by considering the journal, the authors, and the methodology.)

References

- Anson, D. W. J. (2024). The impact of large language models on university students' literacy development: A dialogue with Lea and Street's academic literacies framework. *Higher Education Research & Development*, 43(7), 1465-1478. <https://doi.org/10.1080/07294360.2024.2332259>
- Archila, P. A., Molina, J., & Restrepo, S. (2025). Critically reading science-related texts produced by ChatGPT. *Science & Education*, 34(2), 4627-4662 (2025). <https://doi.org/10.1007/s11191-025-00639-y>
- Athens Technology Center (ATC) & Deutsche Welle. (n.d.). *Truly media*. <https://www.truly.media/>
- Baldrich, K., & Domínguez-Oller, J. C. (2024). The use of ChatGPT in academic writing: A case study in education. Pixel-Bit. *Revista de Medios y Educación*, 71, 141-157. <https://doi.org/10.12795/pixelbit.103527>
- Bohannon, J. (2015). Chocolate with high cocoa content as a weight-loss accelerator. *Global Journal of Medical Research*, 15(K2), 9-14. <https://medicalresearchjournal.org/index.php/GJMR/article/view/922>
- Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative AI as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning*, 19, 479-532. <https://doi.org/10.1007/s11412-024-09437-0>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp0630a>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43). <https://doi.org/10.1186/s41239-023-00411-8>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503-2529. <https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13460>
- Du, J., & Alm, A. (2024). The impact of ChatGPT on English for Academic Purposes (EAP) students' language learning experience: A self-determination theory perspective. *Education Sciences*, 14(7), 726. <https://doi.org/10.3390/educsci14070726>
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, 100(12), 4876-4883. <https://doi.org/10.1021/acs.jchemed.3c00505>

- Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine*, 90, 104512. <https://doi.org/10.1016/j.ebiom.2023.104512>
- Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.
- Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20(2), 1-21. <https://doi.org/10.1007/s40979-024-00149-4>
- Kavadella, A., Dias da Silva, M. A., Kaklamanos, E. G., Stamatopoulos, V., & Giannakopoulos, K. (2024). Evaluation of ChatGPT's real-life implementation in undergraduate dental education: Mixed methods study. *JMIR Medical Education*, 10(1), e51344. <https://doi.org/10.2196/51344>
- Kocatepe, M. (2021). Reconceptualising the notion of finding information: How undergraduate students construct information as they read-to-write in an academic writing class. *Journal of English for Academic Purposes*, 54, 101042. <https://doi.org/10.1016/j.jeap.2021.101042>
- Kostas, A., Paraschou, V., Spanos, D., Tzortzoglou, F., & Sofos, A. (2025). AI and ChatGPT in higher education: Greek students' perceived practices, benefits, and challenges. *Education Sciences*, 15(5), 605. <https://doi.org/10.3390/educsci15050605>
- Liang, W., & Wu, Y. (2024). Exploring the use of ChatGPT to foster EFL learners' critical thinking skills from a post-humanist perspective. *Thinking Skills and Creativity*, 54, 101645. <https://doi.org/10.1016/j.tsc.2024.101645>
- Luke, A. & Dooley, K. (2011). Critical literacy and second language learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning: Volume II*, pp. 856-867, Routledge.
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Marrone, R., Zamecnik, A., Joksimovic, S., Johnson, J., & De Laat, M. (2025). Understanding student perceptions of artificial intelligence as a teammate. *Technology, Knowledge and Learning*, 30, 1847-1869. <https://doi.org/10.1007/s10758-024-09780-z>
- Mendez, J., & Tang, P. (2025). *Student perceptions of achieving intended learning outcomes with generative AI*. [Master's thesis, Lund University]. <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9199497&fileId=9199500>
- OpenAI. (2024/5). *ChatGPT* (Version 3.5) [Large language model]. <https://chat.openai.com/>
- Pitts, G., Marcus, V. & Motamedi, S. (2025). Student perspectives on the benefits and risks of AI in education. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2505.02198>
- Raitskaya, L., & Tikhonova, E. (2025). Enhancing critical thinking skills in ChatGPT-human interaction: A scoping review. *Journal of Language and Education*, 11(2), 5-19. <https://doi.org/10.17323/jle.2025.27387>
- Sari, D. N., Marsella, P. E., & Alfiyan, A. R. (2025). Opportunities, challenges and implications of ChatGPT in the self-directed learning process on the critical thinking skills of management students. *International Journal of Integrative Research*, 3(7), 473-488. <https://doi.org/10.59890/ijir.v3i7.44>
- Šedlbauer, J., Činčera, J., Slavík, M., & Hartlová, A. (2024). Students' reflections on their experience with ChatGPT. *Journal of Computer Assisted Learning*, 40(4), 1526-1534. <https://doi.org/10.1111/jcal.12967>
- Sultan, S., Rofiuddin, A., Nurhadi, N., & Priyatni, E. T. (2017).. The effect of the critical literacy

- approach on pre-service language teachers' critical reading skills. *Eurasian Journal of Educational Research*, 71, 159–174. <https://izlik.org/JA23GR56JD>
- Suriano, R., Plebe, A., Acciai, A., & Fabio, R. A. (2025). Student interaction with ChatGPT can promote complex critical thinking skills. *Learning and Instruction*, 95, 102011. <https://doi.org/10.1016/j.learninstruc.2024.102011>
- Teng, F. M. (2024). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60, e12811. <https://doi.org/10.1111/ejed.12811>
- Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, 15(3), 343. <https://doi.org/10.3390/educsci15030343>
- Wallace, C. (2003). *Critical reading in language education*. Palgrave Macmillan.
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(15). <https://doi.org/10.1186/s41239-024-00448-3>
- Xing, X., He, Z., Xu, H., Wang, X., Wang, R., & Hong, Y. (2024). Evaluating knowledge-based cross-lingual inconsistency in large language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2407.01358>
- Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. *Language Learning in Higher Education*, 14(1), 41-56.

Dr Panagiota Samioti (psamioti@uoc.gr) is a linguist and higher education specialist with extensive experience in academic language teaching, language literacy, and multilingual education. She is Laboratory Teaching Staff at the Writing Centre of the University of Crete and a Fellow of the Higher Education Academy (FHEA, UK). She holds a PhD in Theoretical and Applied Linguistics (University of Crete) and an MPhil in English and Applied Linguistics (University of Cambridge). Her expertise spans academic language pedagogy, digital learning environments, and curriculum design in international higher education contexts. She combines a strong research background in linguistics with over twenty-five years of teaching experience across secondary, tertiary, and adult education sectors in Greece and the UK.