International Journal of TESOL Studies (2026)
Vol. 8 (4) 122-143 https://doi.org/10.58304/ijts.260215

Article

The CLARIFIES+ Framework: A Contextualized Prompting Approach for Generative Artificial Intelligence

Jessie Barrot
National University, Philippines

Received: 17 December 2025 / Received in revised form: 4 February 2026 / Accepted: 12 February 2026/ Available online: 23 February 2026

Abstract
This paper argues that generative artificial intelligence (AI) is becoming an active partner in English language teaching (ELT) for lesson planning, communicative practice, feedback, and assessment support. However, classroom use often depends on ad hoc prompts that yield inconsistent pedagogical fit. Because language tasks are shaped by proficiency, skill targets, genre/register expectations, and sociocultural conditions, ELT needs prompting structures that function as instructional specifications rather than generic technical instructions. In response, the paper proposes CLARIFIES+, a contextualized prompting framework for teaching, learning, and assessment that systematizes ELT-relevant task conditions and boundaries. After reviewing major prompting approaches (e.g., zero-shot, few-shot, chain-of-thought, and contextual prompting) and prominent prompting frameworks, the paper identifies a gap between technical prompt guidance and ELT-centered variables, evaluative alignment, and ethical constraints. CLARIFIES+ addresses this gap through nine core components, namely Context, Limitations, Audience, Role, Intent, Format, Inputs, End product, and Style, plus an adaptive "+" layer for process controls, follow-up clarification, tool use, and safeguards that support accuracy, integrity, and learner safety. The paper illustrates how the framework can support differentiated instruction, learner prompt literacy and self-regulation, and more transparent, consistent AI-supported formative and summative assessment. It concludes by noting limitations, implementation challenges, and future research directions.

Keywords
Contextual prompting, generative AI, prompt engineering, prompting framework, computer assisted language learning

1 Introduction

Generative artificial intelligence (AI) has become an influential technological development that increasingly shapes pedagogical theory, instructional practice, and decision-making in the field of English language teaching (ELT). Teachers and learners use systems such as ChatGPT for lesson design,

*Email: jessiebarrot@yahoo.com

communicative practice, feedback, and assessment support, but outcomes depend strongly on prompt form and contextual detail rather than model capacity alone (Bui & Barrot, 2025; Huang, 2023; Kerr & Kim, 2025). ELT therefore faces a new literacy demand, that is, prompt engineering, or deliberate prompt design that steers model output toward an instructional purpose. Prompt engineering has grown rapidly as a field, with a broad set of techniques and taxonomies that address instruction design, exemplar use, and reasoning scaffolds (Lee & Palmer, 2025; Qian, 2025).

Despite this growth, prompt engineering practice in ELT often relies on improvised prompts and informal heuristics, which yield uneven pedagogical value across tasks and learner groups. This issue matters in language education because task quality depends on proficiency calibration, skill-specific targets, genre and register demands, classroom goals, and sociocultural context. A prompt that lacks these variables may produce fluent text that fails to match learner level, pragmatic norms, or assessment intent. Empirical work on prompt literacy training in EFL contexts also indicates that learners benefit from explicit prompt guidance, while accuracy and authenticity concerns persist and require continued human interaction and oversight (Dillon, 2024). More broadly, research on prompt design by non-experts shows frequent difficulty with articulating constraints, evaluation criteria, and task intent, which undermines reliability and consistency (Zamfirescu-Pereira et al., 2023). The problem warrants attention because ELT settings amplify the consequences of misalignment. Instructional materials, model texts, and feedback can shape learner uptake, motivation, and assessment validity (Heift, 2004; Ryan et al., 2021). Thus, ELT needs prompt structures that support teacher agency and protect learner welfare, not only prompts that elicit fluent content.

Available solutions already exist, but each addresses only part of the ELT need. Technique-level approaches, such as zero-shot, few-shot, and chain-of-thought prompting can improve performance and reasoning depth (Wei et al., 2022). Generative AI-focused ELT guidance also outlines affordances and cautions for classroom use, with emphasis on digital competencies and ethical practice (Barrot, 2023; Law, 2024; Yeo, 2023). These options remain feasible because they require no model retraining and can support rapid classroom iteration. However, these approaches also show limitations for ELT. General prompt methods often treat prompting as transferable technique rather than situated instructional design, so they leave key ELT variables under-specified. Chain-of-thought prompts can increase verbosity and can produce plausible rationales that do not guarantee correctness, so verification still relies on expert judgment (Barrot & Bui, 2026; Turpin et al., 2023). Template approaches can support consistency, but they can omit proficiency differentiation, audience alignment, genre constraints, and safeguards for integrity and learner safety that ELT tasks often require. Evidence from classroom-oriented research also flags accuracy and authenticity issues, even when learners receive prompt training (Dillon, 2024).

This paper addresses these gaps through the CLARIFIES+ framework, a contextualized prompting approach tailored to English language teaching, learning, and assessment. As such, this paper intends to achieve three outcomes. It situates CLARIFIES+ within the current prompt engineering literature and its common approaches and frameworks. It also provides an ELT-centered, ready-to-use structure that encodes context into prompt design. Finally, it advances responsible classroom use through constraints and safeguards that reduce hallucinated claims, inappropriate assistance, and unsafe or unethical outputs.

2 General Prompting Approaches

Prompt engineering has become a central mechanism in shaping interactions between users and large language models. However, the literature reveals that its effectiveness depends on more than technical compliance with prompt formats. Studies consistently indicate that output quality, coherence, and task alignment are influenced less by model architecture alone than by how instructions, examples, and contextual cues are specified (Giray, 2023; Hwang et al., 2025). Prompting approaches primarily differ in the amount and form of guidance they embed in the prompt, which in turn shapes how models infer task

requirements. Zero-shot prompting relies entirely on an instruction, without exemplars or demonstrations (Gupta et al., 2025; Yong et al., 2023). Although this approach has demonstrated simplicity and scalability, its success depends heavily on the implicit alignment between user intent and the model's prior training. Zero-shot prompting often presupposes shared assumptions about task goals, output standards, and evaluation criteria, assumptions that may not hold across contexts. Minimal instruction also frequently results in underspecified task conditions. As a result, zero-shot outputs may appear fluent while failing to satisfy deeper task constraints or quality expectations. The approach therefore privileges efficiency and scalability at the expense of control and interpretability.

Few-shot prompting addresses some limitations of zero-shot prompting by incorporating a small number of task-relevant examples that demonstrate desired output patterns (Lee, U. et al., 2024; Yong et al., 2023). Empirical work has shown that even a limited number of exemplars can substantially improve performance and consistency, particularly for tasks that involve recognizable structural or rhetorical patterns (Kim & Lu, 2024). Few-shot prompting reduces reliance on implicit assumptions by making aspects of the task visible through demonstration. However, this strategy introduces new dependencies. Output quality becomes contingent on exemplar selection, representativeness, and alignment with task goals. Poorly chosen examples may constrain model behavior, amplify bias, or overfit responses to narrow patterns. Few-shot prompting thus shifts the problem from underspecification to potential mis-specification. This requires users to exercise informed judgment in prompt design.

Chain-of-thought prompting further increases instructional explicitness by requesting intermediate reasoning steps before a final response (Lee, A. et al., 2024). Research has shown that this approach improves performance on complex reasoning tasks and enhances transparency by externalizing decision processes (Al Nazi et al., 2025). Studies have also linked reasoning-explicit prompts to greater depth and clarity in evaluative output (Barrot & Bui, 2026). Despite these advantages, chain-of-thought prompting raises analytic concerns. The explicit reasoning produced by models does not necessarily reflect authentic internal processes and may function instead as a post hoc rationalization. Moreover, stepwise reasoning can increase verbosity without guaranteeing relevance or alignment with user intent. While chain-of-thought prompting improves interpretability, it does not eliminate the need for critical evaluation of model output.

Finally, there is contextual prompting which is the practice of embedding explicit, relevant background information in a prompt so an AI model can correctly interpret the user's intent and generate responses that are more precise and context-appropriate (Bali et al., 2026; Giray, 2023). It involves providing detailed context to help the model understand the context or intent of the request and is illustrated through examples showing that specifying factors, such as purpose, focus, constraints (e.g., tone, timeframe, length), and audience helps reduce ambiguity and improves task fit (Bali et al., 2026). Preliminary findings suggest that this type of prompting could have a positive impact on student learning (Liu et al., 2024), assessment (Koenitz et al., 2025), and generative AI performance (Yang et al., 2024). However, the approach warrants a critical reading because "more context" does not automatically equal "better pedagogy" or "more valid assessment."

3 Prompting Frameworks

Prompting frameworks respond to the limitations of general prompting approaches by conceptualizing prompt design as a structured and reflective process rather than a single instruction. Lo's (2023) CLEAR path framework emphasizes concision, logical sequencing, explicitness, adaptability, and reflection. Its strength lies in foregrounding revision and metacognitive awareness, which counters the tendency to treat prompts as static commands. This framework can "systematically resolve" recurring problems such as incoherent or context-inappropriate responses or content (Cain, 2025; Oreški et al., 2025). That claim, however, also marks a key limitation. The article offers persuasive principles and illustrative

examples, but it does not provide an evaluative apparatus that would allow users to test whether "coherent" or "context-appropriate" outcomes have been achieved across tasks or user groups. The Reflective component endorses continuous evaluation and prompt revision, but the discussion locates evaluation primarily in user judgment and informal feedback rather than in explicit criteria, reliability checks, or systematic benchmarking. The Adaptive component also broadens the framework beyond wording and structure to include parameter choices such as temperature, and the paper separately highlights tokens, temperature, and top-p as important technical factors. This inclusion strengthens the framework's realism, but it also exposes an analytic gap. CLEAR treats these parameters as levers that users can "experiment" with. However, it does not specify decision rules, stopping conditions, or risk controls when parameter tuning trades off determinism against novelty. In effect, CLEAR functions as a robust heuristic and instructional scaffold for prompt construction, but its generality shifts the burden of operationalization to the user, especially for goal specification, quality assurance, and evidence-based evaluation (Chen & Langenau, 2026).

The Goal-Prompt-Evaluation-Iteration (GPEI) methodology proposed by Velásquez-Henao et al. (2023) treats prompt engineering as a four-step cycle that moves from goal definition to prompt design, answer evaluation, and iterative revision until the response becomes "adequate." The framework's most consequential move is its insistence that goal definition must precede prompt writing, since the authors argue that the goal should determine prompt structure and should anchor evaluation, and they note that explicit goal definition appears in only one of the methodologies they reviewed. That analytic diagnosis strengthens the framework, as evidenced by recent works (e.g., Barba et al., 2025; Vieira et al., 2025). However, the proposed "systematic evaluation" rests primarily on a short set of diagnostic questions about completeness, accuracy, relevance, compliance with limits, and the presence of hallucinations or contextual inappropriateness. These questions provide a defensible minimal checklist, but they leave open how a user should adjudicate trade-offs among criteria, how much evidence suffices to reject an answer, and what reliability procedures should govern judgments across evaluators or iterations. The framework also imports Explainable AI-inspired evaluation moves, such as requests for reasoning, justification, stepwise breakdown, and inquiry into data sources or training data. This addition signals a commitment to transparency, but it does not resolve the core epistemic problem that plausible justifications can accompany incorrect content, which means evaluation still depends on external verification and human expertise. These limits mirror empirical findings that place human judgment at the center of quality control when users work with AI-generated feedback and revisions (Banihashem et al., 2024; Yan & Zhang, 2024). While GPEI offers a stronger process logic than many generic prompt guides, its prescriptive power remains constrained by under-specified decision rules for goal precision, ethical constraint handling, and evidence standards during iteration.

Giray's (2023) Instruction-Context-Input Data-Output Data framework offers a concrete prompt template through four elements, namely instruction, context, input data, and an output indicator that specifies the desired format. Its main analytic contribution lies in its implicit critique of ad hoc prompting; that is, instructions alone often fail because they omit the background conditions and task-relevant data that allow a model to infer what "good" output should look like, so Giray treats context and input quality as central determinants of relevance rather than optional add-ons (Dornburg & Davin, 2025; Gong et al., 2025). The framework's concreteness, however, also marks its limitation, since it functions primarily as a compositional checklist and does not supply decision rules for how much context suffices, how constraints should be prioritized, or how output quality should be evaluated and corrected across iterations, even as the paper flags risks such as ambiguity, bias reinforcement, lack of context, ethical issues, and unrealistic expectations about model limits. The framework's most distinctive emphasis therefore rests on contextual prompting. Giray defines context as external information that improves relevance and accuracy, and he treats contextual prompts as a distinct technique that steers responses through explicit situational framing, which positions context as a control variable for task fit rather than as decorative prompt detail. Table 1 presents a matrix summary of the reviewed prompt frameworks.

Table 1

Summary of Reviewed Prompt Frameworks

Framework	Core Elements	Strengths	Limitations and Risks
Lo's CLEAR Path (Lo, 2023)	Concise, Logical, Explicit, Adaptive, Reflective	Supports prompt revision and metacognitive control. Encourages explicitness and parameter awareness for output control.	Lacks defined criteria and reliability checks for quality claims. Offers limited decision rules for parameter tuning, which can increase output volatility and teacher workload.
GPEI Cycle (Velásquez-Henao et al., 2023)	Goal → Prompt → Evaluation → Iteration	Enforces goal-first design and aligns prompts with evaluation checks. Provides a minimal checklist for verification routines.	Under-specifies trade-off rules and evidence thresholds across iterations. Explanation requests can yield plausible rationales that still require external verification and domain expertise. Human judgment remains central.
Instruction–Context–Input Data–Output Indicator (Giray, 2023)	Instruction, Context, Input Data, Output Indicator	Offers a concrete template that reduces ad hoc prompting. Elevates context and input specificity; supports format discipline through output indicators.	Functions as a checklist more than an evaluation method. Lacks rules for context sufficiency, constraint prioritization, and correction across iterations. Context overload risk rises without selection rules.

The literature suggests that prompting approaches and frameworks offer complementary but still incomplete solutions. Approaches such as zero-shot, few-shot, and chain-of-thought prompting vary in instructional density, but their effectiveness remains contingent on implicit assumptions, exemplar quality, and the interpretability of model-produced reasoning rather than on guaranteed alignment with complex task conditions. Meanwhile, prompting frameworks provide useful structure and encourage reflection, but they often stop short of specifying domain-sensitive decision rules for how much context is sufficient, how constraints should be prioritized, and how quality should be evaluated and corrected across iterations. This tension underscores the need for a prompting framework that explicitly encode task and context variables, foreground evidence-based evaluation and safeguards, and retain critical human oversight, especially in situated domains, such as ELT where proficiency, genre, register, and sociocultural conditions shape what counts as "good" output.

4 The CLARIFIES+ Framework

The reviewed prompting approaches and frameworks underscore a shared tension between technical guidance and pedagogical specificity. While existing models offer useful heuristics for structuring instructions, examples, and reasoning, they leave largely unarticulated the contextual, instructional, and ethical variables that are central to ELT. In particular, gaps persist in how goals are pedagogically grounded, how learner characteristics and sociocultural conditions are encoded, and how output quality is judged against language-learning purposes rather than generic notions of adequacy or coherence. It is within this space that the proposed CLARIFIES+ framework is situated as a form of contextual prompting tailored to language education. Rather than replacing prior approaches, CLARIFIES+ builds on their insights by systematizing ELT-relevant context into a sequenced prompt structure that makes task conditions explicit. This framework conceptualizes prompting as an instructional design practice rather than a technical interaction with generative AI. The framework responds to the need for systematic

prompt construction that reflects the conditions under which language is taught, learned, and assessed. CLARIFIES+ comprises nine core components, namely *Context*, *Limitations*, *Audience*, *Role*, *Intent*, *Format*, *Inputs*, *End product*, *Style*, and +, which covers optional elements that add process controls and safeguards.

4.1 Context

The Context component establishes the situational frame within which a generative AI system produces a response. It specifies the educational setting, instructional purpose, learner profile, and communicative conditions that shape language use. In ELT, language does not exist in isolation. It emerges from particular social, institutional, and pedagogical circumstances. Context anchors AI output to the realities of classroom practice rather than to abstract or generic language tasks. Within CLARIFIES+, context answers the question of where, why, and under what conditions a task occurs. It may include information about the educational level, proficiency range, instructional modality, curricular focus, assessment purpose, or sociocultural environment. When this information remains absent, generative AI tends to default to generalized academic norms that may misalign with local objectives or learner needs. Explicit context reduces this risk and increases pedagogical relevance. Context also serves a gatekeeping function. It constrains the scope of AI output so that content, tone, and complexity remain appropriate to the specified situation. In communicative language classrooms, such constraint proves essential. Tasks that aim to support interaction, pragmatic competence, or formative assessment require different language choices than tasks designed for test preparation or academic writing. Context ensures that AI responses reflect these distinctions. The following examples contrast a vague prompt with a more specific, pedagogically grounded version.

Vague	Specific
Create a speaking activity for students.	This task takes place in a Grade 10 public school English class in China. Learners demonstrate intermediate proficiency and focus on oral interaction for everyday communication.

4.2 Limitations

The usefulness of a generative AI response often depends less on what the system can produce than on what the prompt forbids. The Limitations component makes that restraint explicit. It tells the system where to stop, what to exclude, and which standards must govern the output. In ELT, such constraints protect pedagogical intent and ethical practice, especially when AI defaults to plausible text that may lack verification, local fit, or appropriateness for learners. Limitations operate as a form of instructional governance. They convert teacher judgment into prompt-level rules. These rules can target factual reliability, source quality, tone, learner safety, or the degree of assistance the tool may provide. When a teacher states, for example, that the system must not invent citations, must rely on peer-reviewed literature, or must avoid sensitive topics, the prompt no longer functions as a request alone. It functions as a bounded task with conditions that mirror professional norms in teaching and assessment. Limitations must remain specific and enforceable. Vague cautionary phrases such as “be accurate” or “be ethical” seldom produce stable results. A prompt benefits more from concrete constraints such as “use only verifiable sources,” “avoid medical or legal advice,” “do not provide answers to test items,” or “offer two alternatives at most.” Such precision makes the output easier to evaluate and revise, and it strengthens teacher agency over AI-supported work. The examples below compare a vague prompt with a clearer, more specific version.

Vague	Specific
Provide a short overview of task-based language teaching and cite key scholars.	Provide a short overview of task-based language teaching. Do not invent citations, publication years, or scholar names. If you remain uncertain, state that uncertainty and omit the claim.

4.3 Audience

The Audience component identifies the intended recipients of a generative AI output and shapes how language, content, and rhetorical choices unfold. Audience awareness remains central to communicative competence. Learners write, speak, and interpret texts in relation to others, not in abstraction. The CLARIFIES+ framework treats audience specification as a core condition for pedagogically meaningful AI output. Audience information guides lexical choice, syntactic complexity, discourse organization, and pragmatic tone. When a prompt fails to identify who the output addresses, generative AI often defaults to a generalized, academically proficient reader. Such default positioning risks mismatches in difficulty, register, and cultural framing, particularly in multilingual and diverse classrooms. Explicit audience specification counters this tendency and aligns AI output with learner realities. Audience also carries implications for instructional equity. Different learner groups require different forms of scaffolding, explanation, and rhetorical support. A text written for secondary school learners differs substantially from one designed for adult professionals or international university students. The audience component allows teachers to encode these distinctions directly into the prompt, rather than relying on post hoc revision. In assessment contexts, audience specification proves equally important. Feedback intended for novice writers should emphasize clarity and encouragement, while feedback for advanced learners may focus on precision, cohesion, or disciplinary conventions. Without a stated audience, AI-generated feedback risks either oversimplification or undue complexity. The following examples illustrate a general prompt and a more detailed alternative.

Vague	Specific
Write an expository paragraph about climate change.	Write an expository paragraph about climate change for first-year university students with intermediate English proficiency.

4.4 Role

Language classrooms rarely treat communication as disembodied text. Utterances carry a speaker identity, an implied authority, and a relationship to the listener. The Role component in CLARIFIES+ captures this social dimension. It tells generative AI who it "is" for the task and, as a result, what kind of discourse it should produce. A role can signal expertise, interpersonal distance, institutional responsibility, or conversational purpose. It can also determine what the system should refuse to do, especially in sensitive interactions. Role matters because generative AI tends to adopt an unmarked expert voice when prompts remain underspecified. That default voice may sound confident, but it may conflict with pedagogy. A teacher often needs a tutor voice that encourages revision rather than a judge voice that evaluates. A speaking practice task often needs an interlocutor voice that negotiates meaning rather than a lecturer voice that explains. Role assignment acts as a switch that selects an interactional frame. Role specification also supports pragmatic competence. Many ELT outcomes require learners to manage politeness, stance, and register across social situations. A role-play that places a learner in front of a "friend," a "supervisor," or a "customer" demands different choices in modality, hedging, and

directness. When generative AI adopts a defined role, it can model those choices and sustain a stable discourse identity across turns, which strengthens the authenticity of classroom simulation. A careful role description can encode pedagogical ethics as well. If a prompt assigns the role of "psychologist" in a chat activity, the prompt must also set boundaries so the system does not offer clinical diagnosis or crisis advice. Thus, role and limitations work together. Role defines the interactional stance, while limitations define the permissible scope of action. Below are examples showing a broad prompt alongside a more precise revision.

Vague	Specific
Have a conversation about stress management.	Take the role of a school counselor. Engage in a short guided conversation with a student to discuss stress management strategies using empathetic and non-judgmental language.

4.5 Intent

The Intent component specifies the purpose that guides the use of generative AI in a given task. It clarifies what the user seeks to accomplish and directs the system toward an appropriate mode of response. Intent distinguishes among tasks that aim to inform, practice, assess, scaffold, or simulate communication. Without such clarification, AI output may drift across functions and dilute instructional value. Intent determines the type of action that the system performs. A request to explain a concept differs fundamentally from a request to model discourse, engage in dialogue, or generate instructional materials. When intent remains implicit, generative AI may combine explanation, evaluation, and creative expansion in ways that conflict with pedagogical goals. The CLARIFIES+ framework therefore treats intent as a central organizing element that stabilizes task direction. Intent also supports alignment across teaching, learning, and assessment. In instructional design, teachers often require concise models rather than extended analysis. In communicative practice, interaction matters more than informational completeness. In assessment contexts, feedback may aim to diagnose rather than correct. Explicit intent communicates these distinctions to the system and reduces unintended output. In multimodal contexts, intent becomes even more critical. Generative AI can produce text, dialogue, images, tables, or structured prompts. A clearly stated intent ensures that the system selects the appropriate output form and level of elaboration. It also clarifies whether creativity, accuracy, or interaction should take priority. The examples that follow present an unclear prompt and its more specific counterpart.

Vague	Specific
Talk about ordering food in a restaurant.	Engage in a role-play conversation to practice ordering food in a restaurant.

4.6 Format

A prompt can specify content with precision and still fail in practice if the output arrives in the wrong shape. The Format component addresses this practical constraint. It tells generative AI what the response should look like, how it should organize information, and which conventions should govern presentation. In ELT, format often determines whether an output functions as a usable classroom resource or as an unfocused block of text. Format choices influence how learners process language. Bullet points support scanning and review. Tables support comparison and classification. Dialogue scripts support interactional practice. A model essay supports genre instruction. Even when the content remains constant, the

format can raise or lower cognitive load and can either scaffold learners or overwhelm them. Format therefore connects directly to pedagogy, not aesthetics alone. Format also communicates disciplinary and institutional expectations. Academic contexts privilege citation and reference conventions, while workplace contexts privilege clarity, concision, and standardized document patterns. A prompt that requires APA references, a rubric layout, or a worksheet template does more than request "nice formatting." It encodes the communicative norms that learners must learn to recognize and reproduce. Format becomes especially consequential in assessment and feedback. Teachers require stable structures such as criterion tables, leveled descriptors, or comment banks because such structures support reliability and transparency. A free-form response may offer useful observations, but it can undermine fairness if learners receive feedback in inconsistent forms. A format constraint guards against that variation. The following examples demonstrate how a vague prompt can be refined into a specific one.

Vague	Specific
Summarize research on vocabulary learning strategies.	Summarize research on vocabulary learning strategies. Present the summary in paragraph form and list references in APA style at the end.

4.7 Input

The Input component refers to the texts, materials, or data that a user provides for generative AI to examine, transform, or respond to. Inputs function as the linguistic and pedagogical basis of AI output. They anchor responses to specific content rather than to generalized assumptions. Without explicit inputs, generative AI relies on probabilistic knowledge that may diverge from instructional goals or classroom materials. Inputs serve multiple pedagogical purposes. They may act as objects of analysis, such as student writing or reading passages. They may function as models, such as sample dialogues or exemplar essays. They may also operate as constraints, such as rubrics, task instructions, or policy documents. Each type of input shapes the direction and boundaries of AI output in distinct ways. The quality and relevance of inputs directly affect output usefulness. A detailed rubric enables feedback aligned with assessment criteria. A well-chosen sample text allows the system to approximate genre conventions and discourse patterns. When inputs remain vague or incomplete, AI output risks misalignment with curricular aims. Clear input specification therefore strengthens coherence between instructional materials and AI-generated responses. Inputs also support transparency and interpretability. When teachers supply the exact materials that guide AI output, learners can trace feedback or models back to shared reference points. This practice reinforces fairness in assessment and clarity in instruction. It also positions generative AI as a responsive tool rather than an autonomous evaluator. Below, a vague prompt is paired with a more targeted, explicit version.

Vague	Specific
Give feedback on this essay.	Use the attached analytic rubric to give feedback on this essay.

4.8 End product

Many prompt failures in ELT do not stem from "wrong" content. They stem from a mismatch between what the teacher expects to receive and what the system decides to deliver. The End product component prevents this mismatch. It names the deliverable in concrete terms, so the task ends with an artifact that fits classroom use. A prompt that asks for "help" or "ideas" leaves too much room for interpretation. A

prompt that specifies an end product makes success observable. End product specification requires more than a topic. It requires a genre or artifact label that tells the system what counts as the final output. In ELT work, the difference matters. A topic such as "climate change" can yield a summary, an expository paragraph, a news report, a debate script, a vocabulary list, or an assessment task. Each output serves a distinct instructional function and reflects different discourse conventions. The end product component selects one of these possibilities and establishes a clear target. This component complements, but does not duplicate, intent and format. Intent states the pedagogical purpose, such as practice, explanation, or assessment support. Format governs layout, such as bullet points, a table, or APA references. End product specifies what the artifact is, such as a model news article, a sample dialogue, a role-play script, a rubric, or an item set. When teachers name the end product explicitly, they reduce unnecessary revision and increase the reliability of AI assistance. End product clarity also supports learner expectations. Model texts and sample dialogues often serve as anchors for performance. Learners interpret them as representations of what the task demands. A well-defined end product therefore supports transparency and reduces confusion, especially when learners must produce comparable outputs. The examples below show a loosely worded prompt and a more clearly defined version.

Vague	Specific
Explain how governments use taxpayers' money.	Write a feature article that explains how governments use taxpayers' money.

4.9 Style

Even when an AI response contains accurate information, it can still fail pedagogically if the voice feels wrong. The Style component addresses this concern. It specifies the tone and linguistic manner that the output should adopt, such as formal, informal, personal, professional, supportive, or authoritative. Style matters in ELT because learners do not only learn vocabulary and grammar. They learn how language enacts relationships, signals respect, and fits institutions. Style shapes meaning through stance. A formal academic voice often relies on cautious claims, technical terms, and impersonal phrasing. An informal peer voice relies on direct address, shorter sentences, and interpersonal cues. A professional voice relies on courteous requests, clear structure, and controlled urgency. A supportive instructional voice relies on encouragement, clarity, and respectful correction. If a prompt leaves style unspecified, generative AI often defaults to a generic "teacherly" register. That register may suit some tasks, but it can undermine others, such as role-play, workplace writing, or feedback that must preserve learner confidence. Style also supports pragmatic development. Learners must choose language that fits social expectations. They must know when to soften a request, how to sound firm without rudeness, and how to express disagreement without face threat. A prompt that sets style explicitly allows generative AI to model these choices consistently. The value lies in the stability of the model. Learners see how tone aligns with situation and how linguistic choices change across contexts. In feedback tasks, style carries ethical weight. Harsh phrasing can discourage learners, while overly flattering phrasing can obscure priorities for revision. Teachers can use the style component to calibrate feedback so it remains constructive, specific, and motivating. The following examples highlight the difference between an imprecise prompt and a specific rewrite.

Vague	Specific
Talk to me about hobbies.	Engage in a short conversation about hobbies using an informal and friendly tone suitable for peer interaction.

4.10 + (Plus)

Finally, the Plus (+) component represents the flexible and adaptive dimension of the CLARIFIES+ framework. While the core elements define the essential structure of a prompt, the Plus component accommodates optional instructions that refine, regulate, or extend generative AI output. It recognizes that effective AI use in English language teaching often requires adjustment beyond an initial prompt, especially when tasks involve complex pedagogy, ethical considerations, or iterative development. The Plus component includes process guidance, follow-up questions or clarifications, tool use, and safeguards that support accuracy, integrity, and learner safety. These elements do not always appear in a prompt at the outset. They often emerge in response to initial output that requires revision, expansion, or correction. The framework treats prompting as a dynamic practice rather than a single-step interaction. Process instructions within the Plus component guide how the system approaches a task. Teachers may require stepwise reasoning, staged output, or explicit justification of choices. Such instructions prove useful in lesson planning, assessment design, and teacher education, where transparency of reasoning matters as much as the final product. Follow-up instructions also form a central feature of the Plus component. Initial AI output may appear broadly useful but misaligned with learner level, task focus, or ethical boundaries. Teachers can issue clarifications that narrow scope, adjust tone, or redirect focus. This iterative exchange mirrors the revision process that characterizes effective teaching practice. Safeguards represent another critical aspect of the Plus component. These safeguards may restrict sensitive content, require uncertainty acknowledgment, or demand alignment with institutional policies. They ensure that generative AI use remains consistent with professional responsibility and learner protection. The following examples show follow-up prompts to further refine the performance of generative AI.

Initial Prompt	Follow-up or Plus Prompt
Write a sample dialogue for making complaints.	Revise the dialogue so it suits lower-intermediate learners and reduces idiomatic expressions.
Summarize research on peer feedback in writing.	If evidence appears mixed or inconclusive, state that explicitly. Avoid claims that lack clear research support.

4.11 Sample prompts

Below are five complete sample prompts with all CLARIFIES+ components explicitly incorporated. Each prompt is written as a single, ready-to-use prompt that a teacher or learner could directly give to a generative AI system.

Scoring an Essay

Components	Instruction
Context	This task takes place in a Grade 11 English class in an urban public school. Learners have intermediate-to-upper-intermediate proficiency and are completing a formative writing assessment on argumentative essays.
Limitations	Do not assign a final grade or score. Do not rewrite the student's essay. Do not invent feedback criteria beyond the provided rubric.
Audience	The feedback is intended for a Grade 11 student writer.
Role	Take the role of an experienced English teacher providing formative feedback.

Intent	The intent is to evaluate the essay's strengths and weaknesses to support revision, not to judge or rank the student.
Format	Present feedback in a table with three columns: Criterion, Strengths, and Suggestions for Improvement.
Input	Use the attached student essay and the attached analytic rubric focusing on thesis clarity, organization, language use, and evidence.
End Product	Produce structured formative feedback aligned with each rubric criterion.
Style	Use a supportive, constructive, and encouraging tone appropriate for adolescent learners.
+ (Plus)	If the essay shows recurring language issues, highlight only two priority areas for improvement. If you are uncertain about any interpretation, state that uncertainty explicitly.

Getting Feedback on a Lesson Prepared by an English Teacher

Components	**Instruction**
Context	This lesson is designed for a Grade 8 ESL class in a multilingual school. Learners have mixed proficiency levels and are learning past tense narrative writing.
Limitations	Do not redesign the entire lesson. Do not introduce new curriculum objectives beyond those stated.
Audience	The feedback is intended for the teacher who designed the lesson.
Role	Take the role of a teacher specializing in English language teaching.
Intent	The intent is to provide reflective, professional feedback to improve lesson effectiveness and learner engagement.
Format	Organize feedback under three headings: Strengths, Areas for Improvement, and Practical Suggestions.
Input	Use the attached lesson plan, including objectives, activities, and assessment tasks.
End Product	Produce actionable feedback that the teacher can apply directly to revise the lesson.
Style	Use a professional, respectful, and collegial tone.
+ (Plus)	If learner differentiation appears insufficient, suggest at most two realistic adaptation strategies. Acknowledge any assumptions you must make.

Chat Conversation with Generative AI

Components	**Instruction**
Context	This task is part of an autonomous speaking practice activity for adult ESL learners practicing everyday conversation outside class.
Limitations	Do not correct every minor language error. Avoid sensitive personal topics.
Audience	The conversation partner is an adult ESL learner with intermediate proficiency.
Role	Take the role of a friendly peer conversation partner.
Intent	The intent is to practice conversational fluency and turn-taking in informal settings.
Format	Conduct a short back-and-forth chat of 8–10 conversational turns.
Input	Use the conversation topic of "weekend activities and hobbies."

End Product	Produce an interactive chat conversation that encourages the learner to respond and continue speaking.
Style	Use an informal, friendly, and natural conversational tone.
+ (Plus)	If the learner gives very short answers, ask gentle follow-up questions to extend the conversation.

Vocabulary Enhancement Practice for Learners

Components	**Instruction**
Context	This activity is for a Grade 6 English class focusing on vocabulary development related to daily routines.
Limitations	Do not include words beyond the CEFR A2 level. Avoid abstract vocabulary.
Audience	The activity is intended for Grade 6 learners with elementary English proficiency.
Role	Take the role of a supportive language tutor.
Intent	The intent is to help learners practice and consolidate new vocabulary through guided use.
Format	Present the activity in three sections: Word List, Practice Sentences, and Short Exercise.
Input	Use the attached target words.
End Product	Produce a short vocabulary practice activity that learners can complete independently.
Style	Use clear, simple, and encouraging language.
+ (Plus)	After the exercise, include an optional challenge question for fast finishers. If instructions may be unclear, rephrase them more simply.

Generating Descriptive Text for a Brochure

Components	**Instruction**
Context	This task supports a tourism unit in a vocational English course for adult learners preparing for hospitality-related jobs. Learners have lower-intermediate proficiency.
Limitations	Avoid exaggerated claims or culturally insensitive descriptions. Do not use idiomatic expressions that may confuse learners.
Audience	The brochure text is intended for international tourists with basic English proficiency.
Role	Take the role of a professional tourism content writer.
Intent	The intent is to model clear and engaging descriptive language suitable for promotional materials about El Nido, Palawan.
Format	Write two short descriptive paragraphs (120-140 words total).
Input	Base the description on the following place: a historic riverside town known for local markets, traditional food, and walking tours.
End Product	Produce sample brochure text that learners can use as a model for their own writing.
Style	Use a friendly, welcoming, and informative tone with simple sentence structures.
+ (Plus)	After the text, briefly list three useful descriptive adjectives from the passage with simple definitions.

A shown below, a concrete prompt-output comparison illustrates what CLARIFIES+ changes in practice when a user moves from an underspecified request to a structured prompt that encodes task goals, learner context, constraints, and evaluation targets. The two sample outputs, as sourced from ChatGPT (OpenAI, 2026), diverge in predictable ways because the prompts encode different task conditions, degrees of constraint, and audience assumptions. The generic prompt leaves destination framing and communicative priorities largely unspecified, so the model defaults to conventional travel-brochure content and rhetoric for El Nido, such as limestone cliffs, lagoons, island-hopping activities, vivid sensory imagery, and high-affect evaluative language. Version 2 encodes CLARIFIES+-style constraints that specify a pedagogical context, a lower-intermediate proficiency target, an intended readership with basic English, a professional writer role, a strict length and format requirement, and explicit safeguards that exclude idioms, exaggerated claims, and culturally insensitive descriptions. These constraints narrow lexical and syntactic range, encourage shorter and more literal sentences, and produce a more instructional and "teachability"-oriented text, especially through the required post-task vocabulary list with simple definitions. The comparison also illustrates a core implication of structured prompting. Constraint specificity can improve alignment with the audience, tone, and classroom usability, while also redirecting content selection away from default destination associations when the prompt stipulates an alternative setting, underscoring the need for teacher verification of factual fit and task validity before instructional deployment.

Version 1: Generic Prompt

Write two short descriptive paragraphs for a brochure about El Nido, Palawan. Describe the location, activities, and experiences visitors can enjoy.

> El Nido, Palawan sits at the northern tip of Palawan Island, where dramatic limestone cliffs rise from crystal-clear turquoise waters. It's a tropical paradise known for hidden lagoons, white-sand beaches, and vibrant marine life. The town itself is laid-back and welcoming, offering stunning views at every turn and sunsets that feel straight out of a postcard.
>
> Visitors can spend their days island-hopping through secret coves, kayaking across calm lagoons, or snorkeling among colorful coral reefs and tropical fish. Adventure seekers can try diving, cliff-jumping, or hiking, while others may prefer slow beach days and fresh seafood by the shore. Whether you're chasing adrenaline or pure relaxation, El Nido delivers an unforgettable blend of natural beauty, adventure, and island serenity.

Version 2: CLARIFIES+ Prompt

This task supports a tourism unit in a vocational English course for adult learners preparing for hospitality-related jobs, and the learners have lower-intermediate proficiency. The brochure text is intended for international tourists with basic English proficiency. The authors should take the role of a professional tourism content writer and model clear and engaging descriptive language suitable for promotional materials about El Nido, Palawan. The output should be two short descriptive paragraphs (120–140 words total). The description should be based on a historic riverside town known for local markets, traditional food, and walking tours. The tone should be friendly, welcoming, and informative, using simple sentence structures. Exaggerated claims and culturally insensitive descriptions should be avoided, and idiomatic expressions that may confuse learners should not be used. After the text, three useful descriptive adjectives from the passage should be listed, each with a simple definition.

El Nido, Palawan welcomes visitors to a small riverside town with a calm pace and a friendly atmosphere. Old streets follow the water, and simple wooden boats pass quietly. The town has a long history, seen in its traditional houses and small public spaces. It is easy to walk around and enjoy the views of the river and nearby hills. Many visitors like the clear layout of the town and the comfortable places to rest.

Walk through local markets to see fresh fish, fruit, and handmade goods. Try traditional food at family stalls, such as grilled seafood and warm rice meals. Guided walking tours are easy to join and help you learn about daily life, local stories, and river traditions. The routes are short, with photo and snack stops, so you can explore at your own pace.

Useful descriptive adjectives

- Riverside: next to a river
- Local: from the area or community
- Traditional: following old customs or ways

5 Theoretical Support and Application of the Framework

The CLARIFIES+ framework draws support from established traditions in language pedagogy, assessment, and learner self-management. In task-based language teaching, task quality depends on explicit specification of communicative purpose, interlocutor relationships, contextual constraints, and criteria for successful performance (Bygate, 2020; Littlewood, 2004). CLARIFIES+ operationalizes these task conditions through Context, Audience, Role, Intent, and Style, which together specify the task's situated communicative demands rather than a decontextualized language product. This structure supports principled task design because it makes visible the conditions under which learners use language, the interactional roles that shape pragmatic choices, and the intended communicative outcome of the activity.

CLARIFIES+ also supports constructive alignment as a practical design logic because it requires an explicit match among intended learning purposes, task conditions, and evaluative expectations. Intent and End product specify what the activity aims to develop and what artifact or performance constitutes evidence of that development. Format and Inputs constrain how evidence appears and which materials or criteria govern interpretation. These components translate the goal-first orientation in prompt design into an instructional specification that makes evaluation targets transparent, which reduces the risk that fluent AI output substitutes for pedagogical fit. This emphasis on explicit criteria and interpretability complements prior prompt-engineering process accounts that foreground goal definition and answer evaluation, while it retains the teacher's evaluative authority over output use in instruction and assessment (Lo, 2023; Velásquez-Henao et al., 2023).

For assessment use, the framework's Limitations, Format, Inputs, and End product components offer prompt-level controls that support validity arguments and reliability-oriented practices in language assessment. Validity requires control of task conditions and defensible interpretation of performance evidence. CLARIFIES+ constrains assistance boundaries, defines permissible output scope, and requires criterion-referenced outputs when teachers supply rubrics or descriptors as Inputs. These design moves reduce construct-irrelevant variance and help preserve assessment purpose, which remains central to responsible language testing practice (Fulcher, 2024). Reliability concerns also benefit from stable response structures. Format requirements, criterion tables, and bounded feedback priorities can reduce uncontrolled variation across AI-generated feedback episodes, a concern that persists in studies that

compare AI-generated feedback with human feedback and learner preferences (Banihashem et al., 2024; Ryan et al., 2021; Steiss et al., 2024).

Finally, CLARIFIES+ supports self-regulated learning by externalizing planning, monitoring, and self-evaluation prompts that learners can adopt as metacognitive routines. Context, Intent, and Limitations require goal setting and boundary awareness; Inputs and End product support monitoring of task completion against explicit criteria; the "+" layer supports revision cycles through clarification requests, uncertainty statements, and staged output. This structure resonates with accounts of self-regulation that emphasize strategic forethought, performance monitoring, and reflective evaluation as core processes that sustain autonomous learning (Zimmerman, 2002). In turn, this theoretical grounding clarifies the framework's intended role as a scaffold for principled teacher decision-making and learner agency rather than as a substitute for professional judgment, especially under known risks in AI-mediated educational practice (Holmes et al., 2022; Luckin et al., 2016).

The integration of a contextualized prompting framework into ELT reflects a broader shift toward principled, transparent, and context-sensitive uses of generative AI. For teachers, they can use CLARIFIES+ as a planning and instructional tool that guides the systematic integration of generative AI into classroom practice. The framework supports lesson design by encouraging teachers to articulate communicative goals, specify linguistic focus, and define contextual constraints before interaction with an AI system. Such explicit specification aligns with research that highlights the value of clear task framing and scaffolding in AI-supported language instruction (Kerr & Kim, 2025; Lee & Palmer, 2025). Through CLARIFIES+, teachers can design prompts that model effective task interpretation, genre awareness, and audience consideration, which remain central to communicative language teaching. CLARIFIES+ also supports differentiated instruction. Teachers can adjust framework components to match learner proficiency, instructional objectives, or curricular demands. For example, prompts can emphasize form-focused feedback for lower-level learners or discourse-level organization for advanced writers, a practice supported by studies on AI-assisted feedback in second language contexts (Wan & Chen, 2024; Yan & Zhang, 2024). The reflective and evaluative elements of the framework further allow teachers to review AI outputs critically, which reinforces teacher authority and pedagogical judgment rather than delegating instructional control to technology.

For language learners, they can apply CLARIFIES+ as a cognitive and metacognitive scaffold for independent learning. The framework supports prompt literacy by guiding learners to clarify objectives, define expectations, and evaluate outputs systematically. Research on learner engagement with generative AI has shown that such structured interaction improves learning outcomes and reduces superficial reliance on AI-generated text (Heung & Chiu, 2025; Zhan & Yan, 2025). Through repeated use of CLARIFIES+, learners develop greater awareness of task demands, linguistic choices, and revision strategies. CLARIFIES+ also supports autonomy by encouraging learners to treat AI as a responsive learning resource rather than an answer provider. Learners can use the framework to request explanations, examples, or alternative formulations that align with personal learning goals. The evaluative and safeguard components of CLARIFIES+ further promote responsible use, as learners must assess output quality and verify alignment with academic norms.

CLARIFIES+ offers practical value for both formative and summative assessment in ELT. Teachers can use the framework to design assessment-related prompts that generate exemplars, feedback models, or diagnostic questions aligned with established criteria. Such alignment supports transparency and consistency, which remain essential principles of valid language assessment. Research on AI-assisted feedback indicates that structured prompt design enhances feedback relevance and learner uptake (Wan & Chen, 2024; Yan & Zhang, 2024). For learners, CLARIFIES+ supports assessment literacy by clarifying expectations and evaluation standards. Learners can use the framework to simulate feedback, compare drafts against criteria, or explore alternative responses prior to submission. This practice supports formative assessment and self-regulation while reducing risks associated with inappropriate AI

use. Ethical considerations also remain central, as the safeguard elements of CLARIFIES+ align with scholarship on academic integrity in AI-mediated writing contexts (Yeo, 2023). Through shared use by teachers and learners, the framework fosters a common evaluative language that strengthens fairness, accountability, and learning-oriented assessment.

CLARIFIES+ differs from, and extends, established instructional design and technology integration frameworks in a domain-specific way. For instance, Willis' (1996) Task-Based Learning framework exemplifies such design logic through a task sequence that specifies pre-task preparation, a task cycle, and a language focus phase, each of which structures pedagogic decisions about task conditions and language attention. CLARIFIES+ does not attempt to replace those models. It specifies the level at which generative AI enters instructional work, namely, the prompt as an instructional specification that governs what the system may produce, for whom, and under which constraints. TPACK also informs this positioning because it conceptualizes effective technology integration as knowledge that connects content, pedagogy, and technology rather than tool use alone (Mishra & Koehler, 2006). CLARIFIES+ complements TPACK through a prompting-specific mechanism that translates that integrated knowledge into explicit constraints and deliverable specifications at the point of AI use. This focus targets a design object that many broader frameworks treat only implicitly, which leaves teachers to translate pedagogical intent into ad hoc prompt language.

The added value of CLARIFIES+ rests on its treatment of prompts as boundary-setting instruments rather than mere input text. Several existing prompt frameworks offer general templates or process cycles, but they do not encode ELT-relevant task conditions with sufficient granularity to function as lesson-ready specifications. CLEAR foregrounds concision, explicitness, and reflective revision, and it highlights the role of parameter awareness, but it leaves evaluation criteria and constraint prioritization largely to user judgment (Lo, 2023). GPEI enforces a goal-first cycle and a minimal evaluation checklist, but it under-specifies decision rules and evidence thresholds that govern safe classroom use and assessment-related boundaries (Velásquez-Henao et al., 2023). Giray's template offers a concrete decomposition of instruction, context, input data, and output indicators, but it functions primarily as a compositional checklist rather than an ELT-centered specification of proficiency, audience, genre, and integrity constraints (Giray, 2023). CLARIFIES+ extends these contributions through systematic encoding of learner-centered variables and pedagogical guardrails that match the realities of language classrooms and assessment settings.

6 Limitations and Challenges

Despite its pedagogical promise, CLARIFIES+ presents several limitations and challenges that require careful consideration. Structured prompting frameworks do not remove the need for professional judgment, nor do they neutralize broader risks associated with generative AI in educational contexts. Research on AI in education consistently emphasizes that instructional quality depends on human oversight, domain expertise, and ethical governance rather than on prompt structure alone (Holmes et al., 2022; Luckin et al., 2016). Thus, CLARIFIES+ should function as a support for decision-making, not as a substitute for pedagogical expertise. One key challenge relates to teacher workload and expertise. Effective use of CLARIFIES+ presupposes familiarity with instructional design, proficiency frameworks, and assessment principles. Teachers who lack confidence in these areas may struggle to specify meaningful constraints or to evaluate AI outputs critically. Studies on educational AI adoption show that uneven teacher preparation often leads to inconsistent or superficial classroom implementation (Ayanwal et al., 2025; Wang et al., 2023). Without sustained professional development, the framework risks reinforcing existing disparities between teachers who possess strong assessment literacy and those who do not.

A second limitation concerns learner dependence and agency. Although CLARIFIES+ encourages explicit goal setting and bounded assistance, learners may still over-rely on AI support if guidance remains insufficiently monitored. Research on self-regulated learning highlights the importance of gradual release of responsibility and reflective practice (Zimmerman, 2002). When prompts provide excessive structure or feedback, teachers and learners may focus on task completion rather than on language development. This risk proves especially salient for novice teachers and learners who may treat AI output as authoritative rather than provisional.

Assessment contexts raise additional challenges. While CLARIFIES+ supports clear boundaries for formative and summative uses, enforcement depends on institutional policy and teacher vigilance. Generative models do not reliably distinguish between permissible guidance and impermissible assistance without explicit constraints, and even well-designed prompts may produce outputs that exceed intended support. Scholarship on language assessment stresses that validity depends on control of task conditions and interpretation of evidence (Fulcher, 2024). Therefore, any AI-mediated assessment practice requires systematic review procedures to confirm alignment with assessment purpose.

Finally, institutional and ethical constraints shape the framework's applicability. Issues of data privacy, transparency, and equitable access remain unresolved across educational systems. Schools and universities that lack clear policies may face uncertainty about acceptable AI practices, which can limit consistent adoption of structured prompting approaches. CLARIFIES+ operates within these systemic conditions and cannot compensate for absent governance or infrastructure. While CLARIFIES+ offers a disciplined approach to prompt design in ELT, its effectiveness depends on teacher expertise, learner guidance, assessment safeguards, and institutional support. Recognition of these limitations remains essential to responsible and pedagogically sound implementation.

7 Future Research Directions

Future research can evaluate CLARIFIES+ through empirical designs that link prompt structure to instructional quality, learner outcomes, and assessment defensibility across contexts. Randomized controlled studies can assign teachers or learners to CLARIFIES+-guided prompting versus ad hoc prompting for parallel ELT tasks such as lesson materials, speaking role-play scripts, and formative feedback. Outcome measures can include expert ratings of pedagogical fit, proficiency calibration, and genre or register alignment, alongside indices of prompt completeness and constraint clarity that reflect documented difficulties in novice prompt design. Research contexts can include secondary EFL classrooms, university ESL writing courses, and teacher-education programs.

Quasi-experimental and design-based studies can embed CLARIFIES+ into sustained classroom routines across several weeks and can examine learner autonomy and self-regulation through repeated cycles of goal specification, monitoring, and reflection. Evidence can include learner self-reports, trace data from prompt revision histories, and teacher judgments of strategic tool use. Mixed-method designs can also examine engagement and participation in AI-supported practice tasks, with attention to sustained use and learner uptake rather than single-session performance.

Assessment-focused studies can test validity and reliability questions that arise when AI supports feedback or rubric-aligned judgments. Comparative studies can examine feedback quality and learner uptake under CLARIFIES+-specified constraints versus common prompting approaches, with analytic attention to prioritization, tone, and criterion alignment. Reliability-oriented outcomes can include inter-rater agreement among human evaluators who judge the usefulness and alignment of AI-generated feedback, as well as stability of feedback structure across repeated prompt runs under fixed Format and Limitations specifications. Validity-oriented outcomes can draw on defensible interpretations of evidence and control of task conditions in relation to the assessment purpose.

8 Conclusion

This paper has articulated the CLARIFIES+ framework as a principled and pedagogically grounded approach to contextualized prompting in AI-supported ELT. The framework synthesizes insights from language assessment, self-regulated learning, and AI literacy to address a central challenge in generative AI use, that is, the gap between technical prompt construction and pedagogical intent. CLARIFIES+ advances prior work on prompt engineering by situating prompts within instructional goals, learner agency, and evaluative alignment rather than treating prompts as isolated technical artifacts. The analysis also positions CLARIFIES+ as a response to emerging concerns related to student engagement, assessment validity, and academic integrity in AI-mediated learning environments. The framework provides a coherent conceptual model that supports responsible, transparent, and educationally meaningful uses of large language models in language-learning contexts, while foregrounding teacher judgment and institutional governance as necessary conditions for pedagogically sound adoption.

References

Al Nazi, Z., Hossain, M., & Al Mamun, F. (2025). Evaluation of open and closed-source LLMs for low-resource language with zero-shot, few-shot, and chain-of-thought prompting. *Natural Language Processing Journal, 10*, 100124. https://doi.org/10.1016/j.nlp.2024.100124

Ayanwale, M. A., Idowu, K. O., Adelana, O. P., Shosanya, S. O., Falebita, O. S., & Adewale, K. A. (2025). Quantifying teachers' readiness for artificial intelligence adoption in education: A mathematical modeling perspective. *Scientific Reports, 15*(1), 26043. https://doi.org/10.1038/s41598-025-08018-x

Bali, A., Kashiramka, S., Guha, A., & Gupta, P. (2026). Contextual and task-specific prompting. In A. Bali, S. Kashiramka, A. Guha, & P. Gupta (Eds.), *Contextual and task-specific prompting* (pp. 33–62). *IGI Global Scientific Publishing*. https://doi.org/10.4018/979-8-3373-0250-8.ch002

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*, 23. https://doi.org/10.1186/s41239-024-00455-4

Barba, G., Lazoi, M., & Lezzi, M. (2025). Enhancing Technological Taxonomies by Large Language Models. In R. Zimmermann, J. Rodrigues, A. Simoes, & G. Dalmarco (Eds.), *Human-Centred Technology Management for a Sustainable Future* (pp. 109–117). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72494-7_12

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Barrot, J. S., & Bui, H. P. (2026). Generative artificial intelligence for automated qualitative feedback: A cross-comparison of prompting strategies. *RELC Journal*. Advance online publication. https://doi.org/10.1177/00336882251412561

Bui, N. M., & Barrot, J. (2025). Using generative artificial intelligence as an automated essay scoring tool: a comparative study. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2521003

Bygate, M. (2020). Some directions for the possible survival of TBLT as a real world project. *Language Teaching*, *53*(3), 275–288. https://doi.org/10.1017/S0261444820000014

Cain, W. (2025). Prompting learning: The EPICC 2 Framework for effective prompt engineering in education. In M. Khine, L. Bognár, & E Afari (Eds.), *Future of Learning with Large Language Models* (pp. 29–58). CRC Press. https://doi.org/10.1201/9781003567257

Chen, H. L., & Langenau, E. (2026). AI prompts for clinical questions: Evaluating information literacy frameworks in health sciences education. *The Journal of Academic Librarianship*, *52*(2), 103217. https://doi.org/10.1016/j.acalib.2026.103217

Dillon, T. (2024). Korean university students' prompt literacy training with ChatGPT: Investigating language learning strategies. *English Teaching, 79*(3), 123–157. https://doi.org/10.15858/engtea.79.3.202409.123

Dornburg, A., & Davin, K. J. (2025). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL*, *37*(3), 332–347. https://doi.org/10.1017/s0958344024000272

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*, 57. https://doi.org/10.1186/s41239-023-00425-2

Fulcher, G. (2024). *Practical language testing* (2nd ed.). Routledge. https://doi.org/10.4324/9781003373629

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering, 51*, 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Gong, X., Xu, W., & Qiao, A. (2025). Exploring undergraduates' computational thinking and human-computer interaction patterns in generative progressive prompt-assisted programming learning. *International Journal of Educational Technology in Higher Education*, *22*(1), 51. https://doi.org/10.1186/s41239-025-00552-y

Gupta, R., Tiwari, S., & Chaudhary, P. (2025). *Generative AI: Techniques, models and applications*. Springer Nature. https://doi.org/10.1007/978-3-031-82062-5_3

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, *16*(2), 416–431. https://doi.org/10.1017/s0958344004001120

Heung, Y. M. E., & Chiu, T. K. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence, 8*, 100361. https://doi.org/10.1016/j.caeai.2025.100361

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promise and implications for teaching and learning*. Center for Curriculum Redesign. https://discovery.ucl.ac.uk/id/eprint/10139722

Huang, J. (2023). Engineering ChatGPT prompts for EFL writing classes. *International Journal of TESOL Studies, 5*(4), 73–79. https://doi.org/10.58304/ijts.20230405

Hwang, M., Lee, K. H., & Lee, H. K. (2025). A word to the wise: Crafting impactful prompts for ChatGPT. *System*, 103756. https://doi.org/10.1016/j.system.2025.103756

Kerr, R. C., & Kim, H. (2025). From prompts to plans: A case study of pre-service EFL teachers' use of generative AI for lesson planning. *English Teaching, 80*(1), 95–118. https://doi.org/10.15858/engtea.80.1.202503.95

Kim, M., & Lu, X. (2024). Exploring the potential of using ChatGPT for rhetorical move-step analysis: The impact of prompt refinement, few-shot learning, and fine-tuning. *Journal of English for Academic Purposes, 71*, 101422. https://doi.org/10.1016/j.jeap.2024.101422

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

Koenitz, H., Barbara, J., & Palosaari Eladhari, M. (2025, December). The GenAI4IDN Benchmark 2.0: AI-assisted assessment of generative AI tools for the design of interactive digital narratives. In M. Reyes & F. Nack (Eds.), *Interactive storytelling* (pp. 254–279). Springer. https://doi.org/10.1007/978-3-032-12408-1_14

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, A., Teo, C., & Tan, S. (2024). Prompt engineering for knowledge creation: Using chain-of-thought to support students' improvable ideas. *AI, 5*(3), 1446–1461. https://doi.org/10.3390/ai5030069

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*, 7. https://doi.org/10.1186/s41239-025-00503-7

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies, 29*, 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

Littlewood, W. (2004). The task-based approach: Some questions and suggestions. *ELT Journal*, *58*(4), 319–326. https://doi.org/10.1093/elt/58.4.319

Liu, H., Zhang, Y., & Jia, J. (2024). The design of guiding and adaptive prompts for intelligent tutoring systems and its effect on students' mathematics learning. *IEEE Transactions on Learning Technologies, 17*, 1353–1363. https://doi.org/10.1109/tlt.2024.3382000

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson. https://discovery.ucl.ac.uk/id/eprint/1475756

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A new framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1177/016146810610800610

OpenAI. (2026). *ChatGPT* (GPT-5.2) [Large language model]. https://chat.openai.com/

Oreški, D., Kišić, A., & Rožman, M. (2025). Prompt engineering as an AI literacy competence: A framework for learners and educators. In *Proceedings of the 16th International Conference on e-Learning* (pp. 160–170). Belgrade Metropolitan University. https://doi.org/10.46793/elearning2025.160o

Qian, Y. (2025). Prompt engineering in education: A systematic review of approaches and educational applications. *Journal of Educational Computing Research, 63*(7–8), 1782–1818. https://doi.org/10.1177/07356331251365189

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021). Designing learner-centred text-based feedback: A rapid review and qualitative synthesis. *Assessment & Evaluation in Higher Education*, *46*(6), 894–912. https://doi.org/10.1080/02602938.2020.1828819

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Turpin, M., Michael, J., Perez, E., & Bowman, S. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems, 36*, 74952–74965. https://proceedings.neurips.cc/paper_files/paper/2023/hash/ed3fea9033a80fea1376299fa7863f4a-Abstract-Conference.html

Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt engineering: A methodology for optimizing interactions with AI-language models in the field of engineering. *Dyna, 90*(SPE230), 9–17. https://doi.org/10.15446/dyna.v90n230.111700

Vieira, C., De la Hoz, J. L., Magana, A. J., & Restrepo, D. (2025). Engineering students' experiences with ChatGPT to generate code for disciplinary programming. *Computer Applications in Engineering Education*, *33*(6), e70090. https://doi.org/10.1002/cae.70090

Wan, T., & Chen, Z. (2024). Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning. *Physical Review Physics Education Research, 20*, 010152. https://doi.org/10.1103/physrevphyseducres.20.010152

Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior, 146*, 107798. https://doi.org/10.1016/j.chb.2023.107798

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*, 24824–24837. https://proceedings.neurips.cc/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

Willis, J. (1996). *A framework for task-based learning*. Longman.

Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications, 11*, 1086. https://doi.org/10.1057/s41599-024-03543-y

Yang, H., Ji, S., Wu, R., & Xu, W. (2024, May). Are you being tracked? Discover the power of zero-shot trajectory tracing with LLMs! In *2024 IEEE Coupling of Sensing & Computing in AIoT Systems (CSCAIoT)* (pp. 13–18). IEEE. https://doi.org/10.1109/cscaiot62585.2024.00008

Yeo, M. A. (2023). Academic integrity in the age of artificial intelligence (AI) authoring apps. *TESOL Journal, 14*(3), e716. https://doi.org/10.1002/tesj.716

Yong, G., Jeon, K., Gil, D., & Lee, G. (2023). Prompt engineering for zero-shot and few-shot defect detection and classification using a visual-language pretrained model. *Computer-Aided Civil and Infrastructure Engineering, 38*(11), 1536–1554. https://doi.org/10.1111/mice.12954

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Article 437, pp. 1–21). ACM. https://doi.org/10.1145/3544548.3581388

Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*. Advance online publication. https://doi.org/10.1080/02602938.2025.2471821

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Jessie S. Barrot (http://orcid.org/0000-0001-8517-4058) is a Professor and Vice President for Research and Development at the National University, Philippines. He has published papers on curriculum and instruction, computer-assisted language learning, language teaching, and L2 writing.