Enhancing collaboration between humans and ChatGPT-3 through effective prompting: Revolutionizing EFL instruction in Moroccan secondary schools.

Amal Zeroauali, Boukare Mohammed

Imane Bakhtaoui, Abdsamad Mimoune Elkhair,

Amal.zerouali@ump.ac.ma

m.boukare@ump.ac.ma

imane.bakhtaoui.d24@ump.ac.ma

abdessamad.mimounelkhir@ump.ac.ma

Communication, Education, digital usages and Creativity.

Abstract

This mixed-methods study examines how Moroccan EFL teachers implement prompt engineering to design lesson plans using Large Language Models (LLMs), particularly ChatGPT, while maintaining alignment with the OHEE scheme. The research methodology combined a quantitative survey (n=249) analyzing teachers' interaction patterns with these tools, followed by qualitative focus group discussions (n=20) conducted in two phases.

The study used purposive sampling to ensure demographic variety and representation across different teaching environments. In the initial phase, teachers without training (n=10) created basic prompts, producing eight first attempts at crafting instructional queries. The second phase involved trained teachers (n=10) in a refined prompt development process, generating 42 prompts across seven conversation-based interactions with ChatGPT. The training introduced a structured framework incorporating five essential elements: persona, audience, context, specific instruction, and output specification.

Quantitative data analysis through SPSS identified usage patterns and behavioral trends, while qualitative analysis examined prompt quality and OHEE scheme alignment. The comparison of 50 prompts (8 basic, 42 refined) showed marked differences between the capabilities of trained and untrained teachers. Results showed that prompts from teachers without training often missed key elements, leading to misaligned educational materials, while trained teachers demonstrated superior ability to create focused, objective-aligned prompts through systematic improvement.

This research advances our understanding of technology-assisted instructional design by demonstrating the value of structured prompt engineering training in educational settings. The findings highlight the importance of systematic professional development in optimizing the use of these tools and point to new directions for research in computer-assisted language teaching methods.

Key words: Large language models, Prompts, Prompt engineering, EFL.

Introduction

The rise of Large Language Models (LLMs) such as ChatGPT has brought transformative potential to the field of applied linguistics, particularly in language teaching and learning. By leveraging advanced machine learning techniques, LLMs are capable of generating human-like texts across a wide range of tasks, offering new opportunities to enhance language education. These models are not only powerful tools for assisting educators but also serve as valuable aids for learners, enabling the creation of personalized and contextually relevant learning materials. Their ability to adapt to different linguistic needs and learning styles marks a significant shift in how language instruction can be delivered.

One of the critical innovations accompanying LLMs is prompt engineering, a process of crafting inputs that guide these models to produce targeted and high-quality outputs. Prompt engineering plays a pivotal role in optimizing the interaction between humans and LLMs, ensuring that the outputs meet specific educational objectives. Through precise prompts, educators can design tailored materials that enhance student engagement and learning outcomes. This capability is particularly valuable in teaching English as a Foreign Language (EFL), where authentic content can support meaningful learning while aligning with their learning goals as specified by the official guidelines.

While previous studies have highlighted the significance of prompt engineering in enhancing AI interactions across diverse educational contexts, research on its application in teacher-centered instructional design remains limited. Most existing literature focuses on student engagement with prompt engineering, exploring its impact on creativity, writing skills, and AI literacy (Bansal, 2024; Ein-Dor, 2024; Slamet, 2024; Velásquez-Henao, 2023; Woo et al., 2023; Hwang et al., 2023). Similarly, studies on EFL teaching emphasize prompt crafting as a tool for generating personalized learning materials but rarely address how educators systematically design prompts to align with specific pedagogical frameworks like the OHEE scheme and specific pedagogical guidelines. Furthermore, the relationship between the quality of prompts and the usability of LLM-generated outputs has not been thoroughly examined, particularly in teacher-led settings. This study seeks to address these gaps by investigating how Moroccan middle school English teachers employ prompt engineering to design lesson activities, evaluating the effectiveness of their prompts, and analyzing how prompt quality influences the accuracy, relevance, and practicality of AI-generated learning materials.

This study aims to explore how English teachers in Morocco utilize prompt engineering to design lesson activities for secondary school students. By analyzing prompts generated during collaborative focus groups, the research seeks to understand how effectively teachers can optimize the potential of LLMs to meet specific instructional goals. Using a conceptual framework that highlights five key elements of prompt engineering—persona, audience, context, specific instruction, and output specification—this study provides insights into the evolving role of AI in language education and the skills educators need to maximize its potential.

Objectives

1. To investigate how English teachers in Morocco use prompt engineering to design lesson activities for middle school students.

2. To evaluate the effectiveness of prompts in generating learning materials that align with the OHEE scheme.

3. To explore the relationship between prompt quality and the accuracy, relevance, and usability of LLM-generated outputs.

Research Questions

1. How do English teachers utilize the five key elements of prompt engineering (persona, audience, context, specific instruction, and output specification) when designing prompts?

2. What is the impact of prompt quality on the relevance and accuracy of LLM-generated learning materials?

3. To what extent do iterative prompt refinements improve the output quality of LLMs in the context of language teaching?

Hypotheses

1. Prompts incorporating all five key elements are more likely to generate accurate and contextually relevant educational content.

2. Teachers who iteratively refine their prompts achieve better alignment between generated outputs and instructional goals.

3. Prompt quality significantly affects the relevance, accuracy and engagement level of LLM-generated EFL learning materials.

1. Concept defining: the interaction between humans and LLMs
2. Large Language Models

Large language models (LLMs) are a subset of generative AI designed to produce human-like language by "iteratively anticipating likely next words based on the sequence of preceding words" (Bommasani et al., 2021; McCoy et al., 2023; Knoth, 2024). These models belong to the broader category of generative AI, which comprises machine learning algorithms capable of learning from diverse content types, including text, images, and audio, to generate new outputs. "The models used in generative AI are capable of producing a variety of outputs, including audio, video, images, or text, based on user input, which is referred to as a prompt" (Knoth et al., 2024, p. 1).

LLMs are foundational to creating conversational AI systems, enabling "smooth and effective multi-turn conversations with users, lowering the barriers to developing conversational user experiences" (Bommasani et al., 2021, cited in Knoth et al, p. 2). Often described as "the brains of AI platforms," these models are built from pre-trained systems that have learned from extensive datasets (Bansal, 2024, p. 15). Their capabilities stem from advances in machine learning, particularly deep learning, and the availability of vast amounts of textual data (Lemeš, 2024, p. 162).

One of the strengths of LLMs is their versatility. They "can process even hybrid sentences in which different languages are mixed" (Lemeš, 2024, p. 161). This makes them invaluable in educational settings, where they can be employed to create "language practice materials for various learning objectives and proficiency" (Kurdi et al., 2020; cited in Hegazy, 2024, p. 507). Their continuous evolution has further enhanced human-machine interaction, making it increasingly natural and efficient (Lemeš, 2024, p. 163).

Given their robust capabilities, LLMs rely heavily on well-constructed inputs to produce high-quality outputs. This brings us to the critical role of prompting, a process central to harnessing the full potential of LLMs. The next section explores the concept of prompts and prompt engineering, outlining how these inputs can shape the quality of AI-generated content.

2.2. Prompt engineering

Prompt engineering is a critical concept in the interaction between humans and generative AI models. At its core, it involves designing, refining, and optimizing specific input to guide large language models (LLMs) like GPT[1] toward producing high-quality outputs (Knoth, 2024; Oppenlaender, 2024). This process not only shapes the AI's response but also establishes the framework for the dialogue by "creating clear guidelines and rules for the LLM's dialogue utilizing a set of predetermined norms" (Knoth et al., 2024, p.2). A High-quality prompt is thus instrumental as it "creates the structure for the dialogue and informs the LLM about which information is important as well as about the intended output form and content" (White et al., 2023, cited in Knoth et al., 2024, p.2).

Prompt engineering, sometimes referred to as "prompting", emerged from the online community of practitioners, initially in the context of creating textual inputs for OpenAI's GPT-3 (Liu & Chilton, 2022; Oppenlaender et al., 2024). It is a multidisciplinary skill, drawing upon fields such as linguistics, logic, philosophy, and engineering to optimize model performance and tailor responses to specific tasks (Lemeš, 2024, pp.165–166). This systematic experimentation, often through trial and error, enables practitioners to refine prompts iteratively, improving the dialogue between humans and AI (Liu & Chilton, 2022; Knoth et al., 2024).

Prompt engineering is more than just a technical process; it is recognized as a skill, defined "as the ability to effectively utilize language and prior knowledge to craft prompts that guide generative models towards desired outputs" (Oppenlaender et al., p.7). This skill involves a "a deep understanding of both the AI model's capabilities and the specific task at hand", requiring practitioners to "avoid ambiguity, providing examples, and iterating carefully" (Bansal, 2024, pp.14–15).

Prompt engineering plays a pivotal role in optimizing model performance, enhancing user experience, and facilitating innovation across various industries (Lemeš, 2024, pp.164–165). It serves as "a bridge between humans and AI" (Bansal, 2024, p.15), enabling more meaningful and efficient interactions. Additionally, it supports customized training and development by tailoring AI responses to specific needs, making it an invaluable tool in

---

[1] GPT refer to pre-trained Generative Transformers and are part of deep learning models. See Velásquez-Henao et al., 2023.

educational contexts, including lesson planning and activity design for middle school students in Morocco (Lemeš, 2024, pp.164–165; Knoth et al., 2024, p.3).

In brief, prompt engineering is a fundamental skill that underpins effective human-AI interaction. By mobilizing strategic prompt construction, educators and other practitioners can harness the full potential of generative AI models, driving innovation and improving task-specific outcomes.

2. Review of literature: Prompt engineering in EFL contexts.

The integration of generative artificial intelligence (AI), particularly LLMs like ChatGPT, has opened new avenues in education, especially in language teaching and learning. In the same line, it has raised many questions and spilled much ink. The following section reviews prior research on prompt engineering, with a focus on its application in EFL contexts, highlighting studies that explore its development, significance, and potential to revolutionize language instruction. The reviewed studies range from those investigating prompt engineering as a literacy skill to those exploring its role in creating personalized educational materials and facilitating teacher-student collaboration.

A study conducted by Woo et al., 2024 with a sample of 27 undergraduate students from a university in Hong Kong aimed to enhance participants' AI self-efficacy, knowledge of generative AI, and prompt engineering skills. Through a 100-minute workshop, the study demonstrated that structured prompt engineering training significantly improved students' ability to craft effective prompts for academic writing tasks. The findings highlight the value of systematic training in transforming intuitive use of generative AI tools into a more purposeful practice.

Hwang et al. (2023) explored the development of "prompt literacy" among 30 EFL students engaged in an AI-powered image creation project. Participants iteratively designed prompts to create artworks reflecting the socio-cultural meanings of English words. The study found that students frequently revised prompts to align outputs with their linguistic and cultural understanding, indicating the potential of prompt engineering in fostering deeper engagement with language learning.

Another study by Woo et al. (2023) examined EFL secondary students' prompt engineering pathways for completing writing tasks with ChatGPT. The research focused on

four case studies, analyzing the trial-and-error processes of students refining their prompts. It found that more sophisticated pathways emerged over time, demonstrating how iterative prompt adjustments could lead to better alignment with task requirements.

A case study by Mena Octavio et al. (2024) followed an experienced EFL teacher in Spain who integrated ChatGPT into her classroom over seven months. Teaching students from A1 to C1 proficiency levels, the teacher refined prompts iteratively to achieve a consistent structure for lesson planning. The study highlighted five essential components—action, format, context, target, and extra information—that enabled ChatGPT to produce coherent and contextually appropriate lesson materials.

Woo et al. (2023) also investigated EFL students' use of generative AI for story writing. With 67 secondary school students in Hong Kong, the study analyzed how participants used prompts to address issues like writer's block and improve their narratives. The research identified three main purposes for prompting: overcoming challenges, developing ideas, and expanding story details, showcasing how prompt engineering can enhance creative writing processes.

Young and Shishido (2024) conducted research on EFL students with a CEFR B2 level to evaluate the impact of prompt engineering techniques on generating educational dialogues using ChatGPT. The study found that clear and detailed prompts significantly improved the quality of AI-generated dialogues for lower-level criteria but faced limitations when generating advanced-level materials. The authors attributed this to ChatGPT's tendency to favor simpler, easily comprehensible responses.

Finally, Lee (2024) explored the use of ChatGPT to design formative assessments for university students engaging with TED talks. The study found that well-crafted prompts, supplemented by iterative refinements, improved ChatGPT's ability to generate meaningful and relevant assessment materials, though human input was often required to finalize the outputs.

The reviewed studies emphasize the transformative potential of prompt engineering in education, particularly in EFL teaching and learning. From improving AI self-efficacy among students to enabling teachers to design contextually relevant lesson plans, the findings illustrate how prompt engineering enhances the interaction between humans and AI. Moreover, these studies highlight the iterative and adaptive nature of prompt crafting, which fosters deeper

engagement and creativity among both learners and educators. This review establishes a foundation for understanding the critical role of prompt engineering in leveraging generative AI tools like ChatGPT, setting the stage for investigating its practical implications for Moroccan EFL teachers. More importantly, the review revealed that research did not address how EFL teachers in Morocco can systematically design prompts that align with the Moroccan English syllabus and the OHEE[2] scheme.

3. Conceptual framework: Prompt engineering and key elements.

The conceptual framework for this study is based on the prompt engineering model developed by Lin (2024). This model identifies five critical elements—persona, audience, context, specific instruction, and output specification—that guide the design and optimization of prompts for large language models (LLMs). These elements provide a structured approach to crafting effective prompts, ensuring that the AI outputs align with specific tasks and user requirements. Figure 1 illustrates these components, which can be task-dependent and may vary based on the instructional objectives.
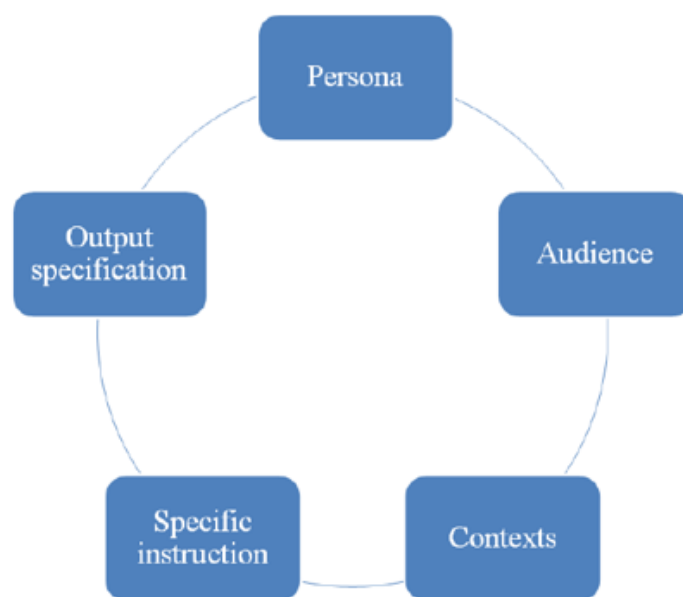

Figure 1: The elements in the prompt. Resource. Source: Lin, Z., 2024, p:15.

Persona (Role):

The persona element refers to the role that the LLM is instructed to adopt within a task. In applied linguistics, LLMs can take on various personas depending on the user's needs. The

---

[2] OHEE is a discovery approach scheme that fosters learner-centeredness. it contains four stages, namely, Observing, Hypothesizing, Experimenting and Extending.

persona might be a virtual teacher for students as it may act as teaching assistants for teachers. It might also include thesis interpreters or research designers (Lin, 2024, p.15). The assigned persona provides the foundation for the LLM's response, aligning its tone, style, and functionality with the user's expectations. In the case of the present study, the persona will be a teacher assistant and material creator.

Audience

The audience defines the intended recipients of the content generated by the LLM. This element is crucial because the suitability of the output depends heavily on the target audience's characteristics. In EFL contexts, the audience can range from kindergarten EFL learners to university-level students, each with distinct language proficiency levels, cognitive abilities, and psychological needs. "These learners have different language abilities, different knowledge backgrounds and different psychological characteristics, which suggest that content generated by LLMs should be tailored to their corresponding characteristics" (Lin, 2024, p.16).

Context

The context element provides the background and situational information needed for the LLM to accurately interpret the task. Contextual details help to ground the AI's response within a specific framework, improving the relevance and coherence of the output. For instance, in a lesson-planning scenario, providing information about the classroom environment, student demographics, or lesson objectives can enhance the AI's ability to generate useful teaching materials (Lin, 2024, p.16).

Specific Instruction

Specific instruction is the cornerstone of prompt engineering. It involves providing clear, detailed, and precise task descriptions to the LLM. The more specific the instructions, the more likely the AI will produce outputs that meet user expectations. Ambiguity in prompts often leads to irrelevant or suboptimal responses, highlighting the need for well-defined directives:

For example, instead of asking, *"Create a lesson plan,"* a specific prompt might request, *"Design a 45-minute lesson plan for intermediate EFL students focusing on past tense verbs, including a warm-up activity, a main exercise, and a homework assignment."* (Lin, 2024, p.16). Despite the importance of specificity, achieving precision can be challenging due to differences in human and machine interpretations of language.

Output Specification

The output specification element outlines the desired characteristics of the AI's response, including its style, tone, format, and structure. This element allows users to tailor the output to their specific needs: For instance, users might request that the output be formatted as a table, bulleted list, or formal document. Output specifications also include parameters for style and tone, such as generating content in a formal academic style or a conversational tone suitable for younger learners (Lin, 2024, p.16).

These five elements form the backbone of effective prompt engineering. By systematically incorporating persona, audience, context, specific instruction, and output specification, users can optimize LLM prompts to generate task-specific and high-quality outputs. This framework is particularly relevant in educational settings, where precise and contextually relevant AI responses can significantly enhance teaching and learning processes.

4. Methodology: Mixed method data collection.

4.1. Design

Figure 2 bellow illustrates the methodology adopted to proceed with the study which adopts a mixed-methods approach to explore how EFL teachers in Morocco utilize prompt engineering for designing lesson plans aligned with the OHEE scheme. The research combines quantitative and qualitative methods to ensure a comprehensive understanding of teachers' interactions with AI tools and the evolution of their prompt-crafting skills. The methodology includes an initial survey to establish usage patterns and perceptions of AI tools, followed by a series of focus group discussions to delve into teachers' practices and evaluate the impact of training on their prompt engineering abilities. Quantitative data were analyzed using SPSS, while qualitative data were subjected to thematic analysis to assess the content and effectiveness of prompts crafted by teachers.
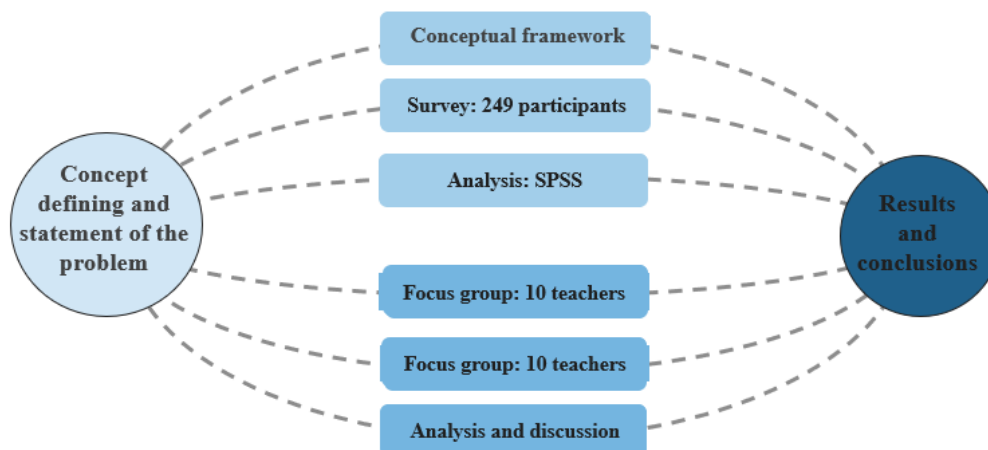
Figure 2: Overview of the research methodology: Investigating prompting practices and lesson design among secondary teachers in Morocco. Source: created by the authors

4.2. Sampling

The study involved EFL teachers from various regions in Morocco, recruited through purposive sampling to ensure diversity in teaching experience and familiarity with AI tools. A total of 249 teachers participated in the survey, providing broad insights into general usage patterns. Two distinct focus groups of ten teachers were subsequently formed to investigate prompt engineering practices in detail. This structured sampling ensured a comparative analysis of untrained and trained teachers' prompt-crafting abilities.

4.3. Survey

The research began with a survey to collect teachers' usage patterns and perceptions of AI tools. It targeted 249 Moroccan EFL teachers. The survey collected quantitative data on teachers' frequency of AI tool usage, behaviors related to prompt crafting, and perceptions of AI-generated lesson materials. These data provided a foundational understanding of how teachers interact with AI tools such as ChatGPT. Analysis using SPSS revealed key trends and informed the design of subsequent focus group discussions by identifying areas requiring deeper qualitative exploration.

4.4. Focus Groups

The qualitative phase consisted of two focus groups, each serving a distinct purpose. The first focus group involved 10 teachers who were tasked with designing lesson prompts without prior training. These participants generated zero-shot prompts[3] -initial attempts at crafting prompts without iterative[4] refinement or structured guidance. This phase yielded eight prompts that were analyzed for alignment with the OHEE scheme and the conceptual framework of prompt engineering.

The second set of focus groups engaged 10 teachers who received targeted training on prompt engineering. This training introduced the five key elements of effective prompt design: persona, audience, context, specific instruction, and output specification. Teachers participated in guided exercises to create iterative prompts through conversation-based interactions with ChatGPT. Over seven conversations, each teacher generated six prompts, resulting in a dataset of 42 prompts for comparative analysis with the zero-shot prompts from the first group.

4.5. Analysis

The study employed both quantitative and qualitative analysis techniques to ensure a thorough evaluation of the data. The survey data were analyzed using SPSS to identify patterns in teachers' AI usage and prompt-related behaviors. For the focus groups, thematic analysis was applied to a sample of 50 prompts to evaluate their quality and alignment with the OHEE scheme. This included eight zero-shot prompts from untrained teachers and 42 iterative prompts from trained teachers. The iterative prompts were analyzed to track improvements over successive interactions, assessing the impact of training on teachers' ability to refine their prompts. Additionally, transcripts of focus group discussions were qualitatively analyzed to capture teachers' reflections, challenges, and strategies in the prompt engineering process. Together, these analyses provided robust insights into the integration of AI tools in EFL lesson design and the evolving role of prompt engineering in educational contexts.

---

[3] The zero-shot approach refers to generating outputs without prior refinement or iterative feedback, relying solely on the initial prompt to guide the AI. (Bansal, 2024; Hegazy, 2024; Knoth et al., 2024; Lin, 2024)

[4] The zero-shot approach refers to generating outputs without prior refinement or iterative feedback, relying solely on the initial prompt to guide the AI. (Knoth et al., 2024; Lin, 2024)

5. Results: Quantitative and qualitative data.

5.1. Quantitative analysis: Patterns of AI usage and prompt behavior

5.1.1. Validity test

Cronbach Alpha	Elements
0,823	9

Table 1: Reliability statistics for survey items on AI usage in teaching

As illustrated in table 1, the reliability statistics, with a Cronbach's Alpha of 0.823 for nine items, indicate good internal consistency of the survey instrument used to evaluate teachers' usage and perceptions of AI tools in teaching. This value, exceeding the commonly accepted threshold of 0.7, suggests that the survey items are well-aligned and effectively measure the intended constructs, such as teachers' behaviors and attitudes toward AI integration. The strong reliability supports the credibility of the survey's findings and provides a solid foundation for interpreting the results with confidence.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
N Valid	249	249	249	249	249	249	249	249	249
Missing	0	0	0	0	0	0	0	0	0
Mean	1,00	4,00	3,82	4,01	3,81	4,22	4,24	4,28	3,89
Median	1,00	4,00	4,00	4,00	4,00	5,00	5,00	4,00	4,00
Standard deviation	,000	,696	,840	,816	1,037	1,032	,901	,767	1,066

Table 2: Descriptive statistics for teachers' usage patterns and perceptions of AI tools in lesson planning and teaching

5.1.2. Survey results

Table 2 provides a quantitative overview of teachers' usage patterns and perceptions of AI tools in lesson planning and teaching. All respondents (N=249) indicated that they have used AI tools for lesson planning (Mean=1/ SD=0), highlighting universal exposure to these tools among the sample.

Teachers generally find AI responses favorable for lesson and material design, as evidenced by a mean score of 4.00 and a median of 4.00 on a 5-point scale, suggesting that most respondents rate the quality as "Good" or higher. Additionally, the mean score of 3.82 indicates that teachers frequently modify their prompts to improve responses. A median of 4.00 suggests that the behavior is common among participants, although some variability exists (SD=0.84).

While modifying the prompts, teachers regularly specify a role for AI in their prompts, as reflected by a mean of 3.81 and a median of 4.00, although this behavior is slightly less consistent compared to other aspects (SD=1.037). Moreover, respondents almost always specify the target audience when using AI, with a mean of 4.22 and a median of 5.00. The highest scores are observed here indicate that most underlining its importance for generating tailored materials.

Similarly, a mean of 4.24 and a median of 5.00 suggest that teachers frequently add situational context in their prompts, reflecting an awareness of how contextual details enhance AI outputs. Gradually, teachers commonly provide detailed instructions in their prompts, with a mean score of 4.28 and a median of 4.00, showcasing an understanding of the relationship between prompt precision and output quality.

While slightly lower than some other metrics, the mean score of 3.89 indicates that specifying the desired output format is a regular practice, with moderate variability among participants (SD=1.066). More importantly, teachers often notice improvements in AI responses following prompt modifications (Mean= 4.01). The consistency (median=4.00, SD=0.816) highlights the importance of iterative querying.

These results demonstrate a strong awareness among teachers of the need for detailed and contextually rich prompts to optimize AI-generated lesson plans. The frequent iterative querying and specification of audience and context reflect an adaptive approach to leveraging AI tools effectively. However, slight variability in behaviors such as role specification and output format suggest areas where training or further exploration could enhance consistency and outcomes.

5.2. Qualitative analysis: evaluating prompt quality and alignment

By focusing on the integration of the OHEE instructional framework and the five elements in Lin's framework, the analysis highlights how training impacts the ability to design lesson plans that promote effective language instruction. The comparison provides insight into

the progression from basic, repetitive task design to more scaffolded, purpose-driven lesson structures that align with modern pedagogical principles.

Prompt	PT	P	A	C	SI	OuS	PQ	Observations
1	OS	ND	D	ND	ND	ND	Low	Audience considered/ lacked specificity, general context and alignment with the adopted scheme
2	OS	ND	D	D	ND	ND	Low	Audience and context considered/ lacked specificity and alignment with the adopted scheme.
3	OS	ND	D	D	ND	ND	Low	Audience considered/ lacked specificity, general context, lacked alignment with the adopted scheme
4	OS	ND	D	ND	ND	ND	Low	Audience considered/ lacked specificity, general context, lacked alignment with the adopted scheme
5	OS	ND	D	D	ND	ND	Low	Audience and context considered/ lacked specificity and alignment with the adopted scheme.
6	OS	ND	D	D	ND	ND	Low	Audience and context considered/ lacked specificity and alignment with the adopted scheme.
7	OS	ND	D	D	ND	ND	Low	Audience and context considered/ lacked specificity and alignment with the adopted scheme.
8	OS	ND	ND	D	ND	ND	Low	Audience and context considered/ lacked specificity and alignment with the adopted scheme.
Prompt type (PT), Persona (P), Audience (A), Context (C), Specific Instruction (SI), Output Specification (OuS), Prompt Quality (QO), Defined (D), Not Defined (ND), One-shot (OS)								

Table 3: Prompt design performance by teachers without training: One-Shot strategy results

Table 3 illustrates the performance of a group of teachers who generated prompts for lesson planning without prior training in the five elements of effective prompt design: Prompt Type, Persona, Audience, Context, and Specific Instruction. These teachers employed a one-shot strategy, producing prompts in a single attempt without opportunities for refinement. Across all prompts, a noticeable lack of specificity and alignment with the adopted teaching scheme was evident. Although the intended audience—8th-grade students in Morocco—was consistently identified, the prompts generally failed to address the learners' prior knowledge, cultural context, or specific instructional needs.

The data reveal significant gaps in the inclusion of key elements. Critical components such as Persona, Specific Instruction, and Output Specification were systematically omitted, leaving the prompts vague and unfocused. Even when lesson topics were specified, as in teaching subject pronouns, prepositions of location, or classroom language, the context and pedagogical approach were insufficiently detailed. This lack of precision likely contributed to the low quality of the resulting outputs, which were uniformly rated as inadequate.

The repetitive shortcomings across prompts indicate a limited understanding of how to effectively structure and tailor instructions to generate meaningful teaching materials. This

issue was further exacerbated by the use of a one-shot strategy, which provided no opportunity to refine or improve prompts based on iterative feedback. Consequently, the outputs were unable to meet the standards of clarity, relevance, or alignment with the adopted curriculum.

Prompt	PT	P	A	C	SI	OuS	QO	Observations
1		D	ND	ND	ND	ND	L	Persona defined/ lacked specificity, general context, lacked alignment with the adopted scheme
2		D	D	ND	ND	ND	M	Persona and audience defined/ lacked specificity and alignment with the adopted scheme.
3	I	D	D	D	D	ND	H	Persona and audience with specified instructions about content and scheme/ no specifications of output.
4		D	D	D	D	ND	H	Persona and audience with specified instructions about content, scheme and time / no specifications of output.
5		D	D	D	D	ND	H	Persona and audience with specified and detailed instructions about content and scheme/ no specifications of output.
6		D	D	D	D	D	V H	Persona and audience with specified instructions about content and scheme and output.
Prompt type (PT), Persona (P), Audience (A), Context (C), Specific Instruction (SI), Output Specification (OuS), Output Quality (QO), Defined (D) Not Defined (ND), Iterative (I). Low (L), Moderate (M), High (H), Very high (VH).								

Table 4: Prompt design performance by trained teachers: Iterative strategy results

Table 4 presents the results of lesson plan prompts designed by a group of teachers who received training in the five elements of effective prompt design. Unlike the first group, this group employed an iterative strategy, refining their prompts to better align with instructional objectives. The analysis reveals notable improvements in the prompts' quality, specificity, and alignment with pedagogical frameworks compared to the first group.

Prompts in this group showed consistent definition of Persona (P) and Audience (A) across all entries, addressing a key weakness observed in the first group. Context (C) and Specific Instruction (SI) were also frequently defined, indicating a better understanding of how to structure prompts to generate coherent lesson plans. For example, in Prompt 3, the four stages of the lesson were clearly outlined, demonstrating a logical flow from observation to production. Similarly, Prompt 4 included precise details such as timing (55 minutes) and the mode of work (pair to group activities), reflecting thoughtful lesson planning.

Several prompts, such as Prompts 3, 4, and 5, detailed a clear framework for the lesson, showcasing how trained teachers incorporated instructional stages and pedagogical strategies. These prompts reached a high (H) quality rating, showing substantial improvement in clarity and alignment with the adopted scheme. Prompt 6, which proposed a detailed table outlining

activities, rationale, timing, and work mode, achieved a very high (VH) quality rating, illustrating the value of defining output specifications.

Despite these improvements, some prompts, such as Prompt 1, still lacked specificity in Context (C) and alignment with the teaching scheme, resulting in a low (L) quality rating. This highlights that even with training, refinement remains essential to address residual weaknesses. However, the iterative approach significantly enhanced the overall quality, as evidenced by the majority of prompts achieving moderate to very high ratings.

Figure 3 and 4 demonstrate examples of lesson plans created by untrained and trained teachers. The lesson plans show notable differences in their approach to structuring activities and aligning with instructional frameworks. Untrained teachers, relied on straightforward, repetitive activities such as vocabulary drills, grammar explanations, and matching exercises. These activities focus primarily on introducing concepts and reinforcing them through basic practice but lack depth in scaffolding or progression. There is minimal engagement with higher-order thinking tasks, such as hypothesizing or applying grammar in creative contexts, which are essential elements of the OHEE scheme. As a result, the plan, as illustrated in Figure 3, remains low-level, emphasizing rote learning over conceptual understanding and application.

Lesson Plan: To Be

Grade Level: 8th Grade

Language Level: Beginner (never studied English)

Topic: Jobs

Grammar Focus: Subject Pronouns & Verb "To Be" (Simple Present Affirmative)

Objectives:

Vocabulary: Students will be able to identify and say common job titles in English.

Grammar: Students will understand and use subject pronouns (I, you, he, she, it, we, they) with the verb "to be" in the affirmative form (am, is, are).

Communication: Students will be able to form basic sentences to describe jobs using "to be" (e.g., "He is a doctor.").

Materials:

Flashcards or pictures of different jobs

Whiteboard and markers

Worksheets with jobs vocabulary and grammar exercises

Audio files (if possible) with pronunciation of job titles

A set of pronoun cards

Lesson Outline:

1. **Warm-Up** (5 minutes)

Activity: Ask students about common jobs they see in their community (teacher, doctor, farmer, etc.), using Arabic or French to prompt them if needed.

Purpose: To activate prior knowledge and introduce the concept of jobs.

2. **Introduction to Vocabulary** (10 minutes)

Activity: Show pictures or flashcards of common jobs (teacher, doctor, farmer, engineer, etc.).

Practice: Say each job aloud, and have students repeat for pronunciation. Write the job titles on the board.

Reinforcement: Point to each picture and ask, "Who is this?" as students respond with the job title, focusing on pronunciation.

3. **Introducing Subject Pronouns** (10 minutes)

Explanation: Briefly explain subject pronouns (I, you, he, she, it, we, they) in simple terms, pointing to yourself for "I," and to a single student for "you," etc.

Practice: Use a mix of pronoun and job cards. For example, hold up "he" and "doctor" and say "He is a doctor." Have students repeat and practice with different pronouns and jobs.

4. **Teaching the Verb "To Be" in Simple Present** (10 minutes)

Explanation: Write the forms of the verb "to be" on the board (am, is, are) and explain their use with each pronoun:

I am; You are; He/She/It is; We are; They are

Examples: Write sentences on the board (e.g., "I am a teacher," "She is a doctor," "They are engineers") and have students repeat each one.

5. **Guided Practice** (10 minutes)

Activity: Pair students up and give each pair a set of pronoun and job cards.

Practice: Students will form sentences using a pronoun and a job card, such as "She is a teacher" or "They are doctors."

Feedback: Monitor pairs and provide feedback on pronunciation and sentence structure.

6. **Independent Practice** (10 minutes)

Worksheet: Distribute a worksheet with fill-in-the-blank sentences (e.g., "___ is a farmer. (He/She/They)"), simple matching activities, and sentence construction tasks.

Purpose: Reinforce understanding of subject pronouns and the verb "to be" with job vocabulary.

7. **Closing Activity** (5 minutes)

Group Review: Go over answers from the worksheet together. Ask each student to say a sentence about one job (e.g., "He is a doctor").

Figure 3: Lesson plan created by untrained teachers using a One-Shot strategy

Conversely, the lesson plan by trained teachers demonstrates a more structured and detailed approach, incorporating clear feedback mechanisms and progression in guided practice. While still not fully aligned with OHÉE, this plan shows a stronger emphasis on experimentation, with activities that scaffold learning through structured pair work and worksheets. Although the plan lacks explicit hypothesizing or creative, real-life application tasks, it includes more varied and purposeful practice activities. This improvement highlights the impact of training on enhancing lesson design, moving from a focus on basic repetition to a more thoughtful integration of instructional elements.

The analysis reveals a clear distinction between the lesson plans produced by the two groups. Untrained teachers tend to design plans that prioritize basic repetition and lack alignment with comprehensive frameworks such as OHEE. In contrast, trained teachers

Stage	Timing	Mode of Work	Activities	Rational/ objective
1. Observing	10 minutes	Group Work	- Students read a short text about vocational education.	- Understand the past perfect tense in context.
			- Text example: *Hassan had always been interested in mechanics. After he had finished high school, he decided to join a training center.*	
			- Groups answer comprehension questions:	
			1. *What had Hassan done before starting his first class?*	
			2. *Why was Hassan successful during his training?*	
			- Teacher elicits sentences using past perfect from the text and explains its use.	
2. Hypothesizing	15 minutes	Pair Work	- **Receptive Task:** Compare two sentences written on the board:	- Distinguish between past perfect and past simple tenses.
			1. *Hassan had learned some skills before he started his training.*	
			2. *Hassan learned some skills, and then he started his training.*	
			- Pairs identify the sequence of events and discuss the difference.	
			- **Productive Task:** Pairs complete sentences using the correct verb form:	
			1. *By the time Fatima _____ (apply) for the job, she _____ (finish) her training.*	
			2. *After the students _____ (complete) their internship, they _____ (receive) their certificates.*	
3. Experimenting	15 minutes	Pair Work & Group Work	- **Speaking Activity:** Pairs discuss their personal experiences using prompts (e.g., *What had you learned before starting baccalaureate?*).	- Practice using the past perfect tense in spoken language.
			- **Group Writing Task:** Groups write a short paragraph about a school or vocational experience, using the past perfect tense.	
4. Extending	10 minutes	Homework	- **Real-Life Application:** Students write a paragraph (6-8 sentences) about a vocational or educational experience.	- Use the past perfect tense to describe real-life situations.
			- Example prompt: *Describe a day when you achieved something important in your studies or training.*	

Figure 4: Lesson plan designed by trained teachers incorporating structured iterative strategy

demonstrate a more structured approach, incorporating iterative feedback and scaffolding to enhance instructional quality. While both plans have room for improvement, the findings underscore the importance of training in equipping teachers with the skills necessary to create pedagogically sound and engaging lesson plans. These results contribute to a broader understanding of how targeted professional development can optimize the use of AI tools in lesson planning.

6. Discussion

6.1. The five key elements and prompt engineering in English lesson design

In relation to the first research question, the results reveal a significant disparity in how the five elements are utilized by trained versus untrained teachers. Teachers without training in

prompt engineering consistently failed to incorporate critical elements such as Persona, Specific Instruction, and Output Specification, resulting in vague and unfocused prompts. While the Audience was routinely identified (e.g., 8th-grade students in Morocco), these prompts lacked adequate contextualization of learners' cultural and educational backgrounds or specific instructional needs.

In contrast, trained teachers displayed a marked improvement in utilizing all five elements. For instance, their prompts frequently included Persona (e.g., identifying the AI as a teaching assistant), Audience, and Context (e.g., specifying lesson timing and mode of interaction). Additionally, trained teachers demonstrated a more comprehensive approach by incorporating Specific Instruction and detailed Output Specifications to guide the AI in generating lesson plans aligned with the adopted curriculum. These findings align with Mena Octavio et al. (2024), who emphasized the role of well-defined components in producing coherent and contextually relevant educational materials.

The significant disparity in how trained and untrained teachers utilized the five key elements of prompt engineering aligns with findings in existing literature. The omission of critical elements such as Persona, Specific Instruction, and Output Specification by untrained teachers mirrors challenges highlighted in Hwang et al. (2023), who emphasized that learners often struggle to craft precise and contextually rich prompts without adequate guidance. Their study revealed that iterative drafting and refining prompts were critical in aligning outputs with socio-cultural and educational contexts. Similarly, Mena Octavio et al. (2024) demonstrated that the integration of structured prompt elements—such as action, context, and target—enabled teachers to achieve coherent and high-quality outputs. This suggests that training fosters the ability to contextualize prompts effectively, aligning them with specific educational objectives and audience needs

6.2. Prompt quality and the relevance and accuracy of LLM-generated learning materials

Prompt quality directly influenced the relevance and accuracy of AI-generated outputs. Prompts from untrained teachers, which often omitted critical elements, led to outputs rated as low quality due to their misalignment with pedagogical goals and insufficient contextualization. For instance, prompts addressing basic topics such as subject pronouns or classroom language failed to detail lesson stages, scaffolding, or alignment with instructional frameworks like the OHÉE scheme.

Conversely, trained teachers produced high-quality prompts that resulted in more relevant and accurate learning materials. Prompts like those specifying detailed lesson plans, instructional strategies, and timing achieved high or very high ratings, demonstrating how prompt quality impacts the ability of LLMs to generate contextually appropriate and pedagogically sound outputs. These results confirm Hypothesis 1, supporting the claim that incorporating all five elements enhances the accuracy and contextual relevance of educational content.

The impact of prompt quality on the relevance and accuracy of AI-generated learning materials is strongly supported by prior studies. Lee (2024) found that prompts with detailed context and instructions generated more meaningful and accurate responses from AI, although human refinement was still required to finalize outputs. This resonates with the current study's findings, where trained teachers' prompts, enriched with specific instructions and audience context, consistently produced outputs that were relevant and pedagogically aligned. Additionally, Woo et al. (2024) demonstrated that structured training in prompt engineering transformed intuitive practices into systematic and effective approaches, enhancing the quality of AI-generated academic materials. These parallels validate the hypothesis that prompt quality is critical for generating accurate, contextually appropriate outputs that meet instructional goals.

6.3. Iterative prompt refinements and prompt quality

The iterative approach adopted by trained teachers significantly improved output quality. Unlike the one-shot strategy of untrained teachers, iterative refinements allowed trained teachers to adjust their prompts based on feedback, ensuring better alignment with instructional objectives. For example, one teacher's prompt underwent several iterations to specify the lesson structure, timing, and collaborative work modes, resulting in a very high-quality output.

This aligns with Woo et al. (2023) and Hwang et al. (2023), who found that iterative prompting fosters improved alignment and task relevance. Furthermore, the data support Hypothesis 2, illustrating how refinement enables teachers to address residual weaknesses and produce higher-order materials that meet pedagogical standards

The results strongly support the hypothesis that prompts incorporating all five elements—Persona, Audience, Context, Specific Instruction, and Output Specification—are more likely to generate accurate and contextually relevant educational content. This was evident in the outputs of trained teachers, whose prompts integrated these elements and consistently produced superior learning materials.

The findings also confirm that iterative prompt refinements improve alignment between generated outputs and instructional goals. The trained teachers' ability to refine prompts iteratively led to better-structured lesson plans, showcasing the value of feedback in achieving pedagogical precision.

In relation to the third hypothesis, prompt quality significantly affects the relevance, accuracy, and engagement level of LLM-generated ESL learning materials. Trained teachers' prompts, which included detailed instructional guidance and cultural context, resulted in outputs that were more engaging and aligned with learners' needs compared to the repetitive and simplistic outputs generated from untrained teachers' prompts. That is the results confirm the third hypothesis as well.

Conclusion

This study has investigated the transformative role of prompt engineering in integrating Large Language Models (LLMs) into EFL teaching, highlighting the critical need for teacher training in crafting effective prompts. By analyzing Moroccan middle school English teachers' use of prompt engineering, the research demonstrates that incorporating the five key elements— persona, audience, context, specific instruction, and output specification—significantly enhances the quality, relevance, and usability of AI-generated learning materials. Furthermore, iterative refinements in prompt design were shown to improve output alignment with instructional objectives, supporting the hypothesis that prompt quality directly impacts educational outcomes. These findings emphasize the potential of LLMs to create contextually tailored, engaging, and pedagogically sound materials when guided by well-structured prompts.

To optimize the full potential of LLMs in language education, it is recommended that teacher training programs incorporate prompt engineering workshops, emphasizing the five key elements and iterative design techniques. Professional development initiatives should focus on practical applications of LLMs in lesson planning, aligning outputs with pedagogical frameworks like the OHEE scheme. Additionally, institutions should encourage collaborative efforts among educators to share best practices and refine prompt engineering strategies. By equipping teachers with these skills, educational systems can foster more personalized and effective language instruction, leveraging AI to address diverse learner needs and enhance teaching outcomes.

However, one limitation of this study is that teachers were not provided with a fully developed prompt incorporating all five key elements (persona, audience, context, specific instruction, and output specification) during the zero-shot phase. This may have constrained their understanding of optimal prompt design from the outset, potentially affecting the quality and effectiveness of their initial outputs and subsequent iterations.

Future research could explore pathways to enhance prompt engineering skills among EFL teachers through longitudinal studies, examining how sustained training impacts their ability to design pedagogically effective prompts over time. Additionally, investigating the integration of prompt engineering frameworks into teacher education programs could provide insights into preparing educators to leverage AI tools effectively. Comparative studies across diverse educational contexts, focusing on different pedagogical frameworks or language proficiency levels, could also offer a broader understanding of how AI-driven lesson design aligns with various teaching schemes. Finally, incorporating student feedback on AI-generated materials could assess the practical effectiveness of prompts in fostering meaningful learning experiences.

Bibliography

- Bansal, P. (2024). Prompt engineering importance and applicability with generative AI. *Journal of Computer and Communications, 12*, 14–23. https://doi.org/10.4236/jcc.2024.1210002
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, et al. (2021). On the opportunities and risks of foundation models. *arXiv*. https://arxiv.org/pdf/2108.07258
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., et al. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv*. https://arxiv.org/pdf/2303.04226
- Ein-Dor, L., Toledo-Ronen, O., Spector, A., Gretz, S., Dankin, L., Halfon, A., Katz, Y., & Slonim, N. (2024). Conversational prompt engineering. *arXiv*. https://arxiv.org/abs/2408.04560
- Hegazy, H. M. M. (2024). A linguistically developed prompt engineering parameters model for enhancing AI's generation of customized ESL reading texts. *Faculty of Education Journal - Alexandria University, 34*(3), 501–566.

- Hwang, Y., Lee, J. H., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of a new literacy skill using generative AI. *arXiv*. https://doi.org/10.48550/arXiv.2311.05373

- 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*, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

- Korzynski, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review, 11*(3), 25–37. https://doi.org/10.15678/EBER.2023.110302

- Lee, D. (2024). Using ChatGPT to design assessments to encourage HE students to engage with TED talks as supplementary course readings: A prompt engineering experiment. In *Proceedings of the Higher Education Research and Development Society of Australasia (HERDSA) Annual Conference* (pp. 8–11). University of Adelaide.

- Lemeš, S. (2024). Prompt engineering. In *Proceedings of the Artificial Intelligence in Industry 4.0: The Future That Comes True Conference* (pp. 159–170). https://doi.org/10.5644/PI2024.215.08

- Lin, Z. (2024). Prompt engineering for applied linguistics: Elements, examples, techniques, and strategies. *English Language Teaching, 17*(9), 14–25. https://doi.org/10.5539/elt.v17n9p14

- Liu, V., & Chilton, L. B. (2023). Design guidelines for prompt engineering text-to-image generative models. *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, April 29–May 5, New Orleans, LA, USA. *arXiv:2109.06977v3 [cs.HC]*. https://doi.org/10.1145/3491102.3502086.

- Mena Octavio, M., González Argüello, V., & Pujolà, J.-T. (2024). ChatGPT as an AI L2 teaching support: A case study of an EFL teacher. *Technology in Language Teaching & Learning, 6*(1), 1–25. https://doi.org/10.29140/tltl.v6n1.1142

- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2024). Prompting AI art: An investigation into the creative skill of prompt engineering. *arXiv*. https://arxiv.org/abs/2303.13534

- Shin, J., Tang, C., Mohati, T., Nayebi, M., Wang, S., & Hemmati, H. (2023). Prompt engineering or fine tuning: An empirical assessment of large language models in

automated software engineering tasks. *arXiv*.
https://doi.org/10.48550/arXiv.2310.10508

- Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence, 4*(46). https://doi.org/10.1007/s44163-024-00143-2

- 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*(230), 9–17. https://doi.org/10.15446/dyna.v90n230.111700

- Woo, D. J., Guo, K., & Susanto, H. (2023). Cases of EFL secondary students' prompt engineering pathways to complete a writing task with ChatGPT. *arXiv*. https://doi.org/10.48550/arXiv.2307.05493

- Woo, D. J., Guo, K., & Susanto, H. (2023). Exploring EFL students' prompt engineering in human-AI story writing: An Activity Theory perspective. *arXiv*. https://doi.org/10.48550/arXiv.2306.01798

- Woo, D. J., Wang, D., Yung, T., & Guo, K. (2024). Effects of a prompt engineering intervention on undergraduate students' AI self-efficacy, AI knowledge, and prompt engineering ability: A mixed methods study [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.32577.36968

- Young, J. C., & Shishido, M. (2024). Prompting brilliance: Unlocking ChatGPT's potential to revolutionize EFL dialogue practices. In *Proceedings of the 30th Annual Conference of the Japanese Association for Natural Language Processing* (pp. 3155–3159). https://creativecommons.org/licenses/by/4.0/