Prompt Engineering Frameworks and Their Educational Value in Higher Education

Siti Khadijah Mohamad[1*], Harvinder Kaur Dharam Singh [1]
[1]Faculty of Education, Open University Malaysia, Malaysia
**Corresponding author: Siti Khadijah Mohamad (siti_khadijah@oum.edu.my)*

Received: 30 November 2025
Received in revised form: 14 December 2025
Accepted: 16 December 2025
Published: 18 December 2025

ABSTRACT

Generative artificial intelligence (GenAI) is reshaping how students and educators access, construct, and evaluate knowledge. Central to this interaction is the prompt, which is the structured natural language input that steers large language models (LLMs). While prompt engineering is increasingly discussed in higher education, there is limited synthesis of how concrete prompt engineering frameworks and prompt patterns add educational value. This review analyses five key studies that offer explicit structures for prompting. These include a prompt pattern catalogue for LLM-based software development, a prompt pattern sequence for software architecture decision-making, an empirical study of student translators' prompting behaviours, an automatic question generation system grounded in a teacher knowledge base, and an MCQ generation framework that integrates retrieval-augmented generation with chain-of-thought and self-refine prompting. The paper is guided by a single research question that examines what forms of educational value these frameworks demonstrate or imply for teaching, assessment, and curriculum design. The integrative analysis shows that such frameworks scaffold the teaching of prompt design, support AI-assisted assessment while maintaining quality, and position prompt engineering as a transversal curriculum competence. At the same time, evidence of intuitive and uninformed student prompting underscores the need for explicit, structured instruction and further research on assessing prompt engineering skills.

Keywords
Prompt engineering; Generative artificial intelligence; Prompt framework; Higher education

Introduction

The uptake of generative artificial intelligence (GenAI) has altered the texture of learning and teaching in higher education. Text-based large language models (LLMs) can generate explanations, code, feedback, and assessment items with a few lines of natural language input (Chang et al., 2024). For many students, interacting with an LLM has become as routine as searching the web. For many educators, these systems represent both an opportunity and a concern because they can support learning while simultaneously complicating questions of academic integrity, authorship, and assessment (Perkins, 2023).

At the heart of this interaction lies the prompt. A prompt is not simply a casual question. It is a structured set of instructions that strongly shapes the behaviour of the model. White et al. (2023) describe prompts as "a set of instructions" that can effectively programme an LLM by customising and refining its capabilities, influencing not just a single answer but the ongoing dialogue and the form of generated output (e.g. code, tables, stepwise reasoning). In practice, prompts can specify roles ("You are an experienced software architect…"), embed context (learning outcomes, constraints), define the output format (rubrics, test items), and shape interaction patterns (ask questions first, then produce a solution).

In educational settings, the quality of a prompt can make the difference between shallow or misleading outputs and responses that genuinely support understanding, critical reflection, or task completion (Geroimenko, 2025). Work in automatic question generation and multiple-choice question design demonstrates that carefully structured prompts, often combined with external knowledge sources, can produce assessment items that meet educational standards and approximate the quality of manually authored questions. Yet empirical work on student practice suggests that many learners use GenAI in intuitive ways. In translator education, for example, Zhang et al. (2025) found that postgraduate students rely heavily on GenAI tools but often use prompts that are one-shot, informal, and lacking in context, leaving much of their disciplinary knowledge unexpressed in their interactions with the model.

At the same time, there is a growing body of work that attempts to move beyond ad hoc advice ("be specific", "add more detail") towards frameworks and pattern catalogues for prompt engineering. Some of these frameworks originate in software engineering and software architecture; others are developed directly for educational tasks such as question generation and personalised learning. What they share is an attempt to formalise effective ways of interacting with LLMs, documenting recurring solutions as patterns, sequences, or workflows.

These developments suggest an important shift. If prompts can be designed systematically, and if there are reusable patterns for structuring them, then prompt engineering becomes a plausible object for teaching, assessment, and curriculum design (Lee & Palmer, 2025). Instead of treating GenAI use as an unstructured risk, higher education could ask: *What exactly should students know about prompt design? How can they be taught to construct, refine, and critique prompts in ways that support learning outcomes? And what role can structured frameworks play in that process?*

This paper addresses these questions by focusing on a single research question:
What forms of educational value do existing prompt engineering frameworks and prompt patterns demonstrate or imply for teaching, assessment, and curriculum design in higher education?

To answer this, the paper reviews five key studies that propose or operationalise prompt engineering frameworks with clear educational relevance. The review does not aim to be exhaustive. Rather, it offers an examination of how these frameworks work and what they afford, with a particular focus on the design of learning, assessment, and curricula in higher education.

Literature Review

GenAI Prompt Engineering and GenAI in Higher Education

Prompt engineering is often defined as the process of writing, refining, and optimising prompts to obtain responses from GenAI systems that are accurate, relevant, and aligned with user intent (Schulhoff et al., 2025). It involves selecting appropriate roles, specifying context, constraining output formats, and iteratively adjusting the input in response to the model's behaviour. Conceptual work on GenAI in education positions prompt engineering as a crucial aspect of AI literacy, as users without an understanding of how to structure prompts are more likely to obtain superficial, biased, or misleading outputs (Xu et al., 2024).

Empirical studies of GenAI use in higher education show that staff and students are experimenting with prompts for a variety of purposes, including generating lesson plans, drafting explanations, exploring alternative assessment tasks, and supporting creative work. In many of these cases, educators describe moving from simple, single-turn prompts towards more deliberate strategies such as defining personas, feeding curriculum documents into the model, or iteratively refining prompts based on outputs. However, this practice is often described in narrative terms rather than grounded in explicit frameworks.

The challenge for higher education is thus twofold. First, students are already using GenAI tools in formal and informal learning, as Zhang et al.'s (2025) translation study illustrates, yet their interactions often lack structure and critical reflection. Second, while educators are developing richer prompting practices, these often remain tacit. Frameworks and pattern catalogues offer one way to make these practices explicit, potentially enabling more systematic teaching, assessment, and curriculum integration.

Prompt Engineering Frameworks and Prompt Patterns

The idea of using patterns to document recurring solutions has a long history in software design. White et al. (2023) extend this tradition to GenAI by introducing a prompt pattern catalogue. Each prompt pattern is documented in a standardised form that includes a name, classification, intent, context, motivation, structure, example implementation, and consequences. The patterns cover a range of categories, such as output customisation (e.g. asking for tables, code, or specific formats), error identification and correction, prompt improvement, interaction management, and context control. A pattern might, for instance, describe how to:

- ask the model to reason step by step before providing an answer
- specify a particular persona ("senior developer", "examiner") to influence tone and content
- request that the model critique its own output before revision

The crucial point is that these are not ad hoc tricks but documented ideas that can be named, taught, and recombined. Maranhão and Guerra (2024) build on this concept by developing a prompt pattern sequence for software architecture decision-making. Their sequence includes patterns such as *Software Architect Persona*, *Architectural Project Context*, *Quality Attribute Question*, *Technical Premises*, and *Uncertain Requirement Statement*. Each pattern is described using an extended pattern form inspired by Coplien's (1998) work on design patterns, including elements such as statement templates, concrete examples, related patterns, usage examples, and known uses.

What distinguishes this contribution is the emphasis on sequence. Rather than treating patterns as standalone tools, the Maranhão and Guerra (2024) show how a series of prompts can progressively establish a persona, define context, surface uncertainties, and ask the model to explore trade-offs. Applied to real industrial cases, the sequence demonstrates that LLMs can support complex architectural reasoning when guided by well-structured prompts, while still requiring human judgement and oversight. Both papers illustrate how prompt engineering can be formalised as a pattern-based practice, suggesting clear parallels with the way design patterns are used in software engineering education.

Prompting in Translation and Assessment Contexts

Beyond software-focused work, prompt engineering is increasingly explored in language and assessment domains. In translator education, Zhang et al. (2025) investigate how student translators actually use GenAI tools over an eight-month period. They collect real prompts used by students in translation and revision tasks and analyse them qualitatively. The study reports that students tend to use short, informal prompts, often phrased as single-round queries. Contextual information about genre, intended audience, or quality requirements is frequently omitted. Yet students do embed some theoretical knowledge, such as asking the model to explain certain translation strategies or to provide alternative expressions. Zhang et al. (2025) argue that such findings can inform specialised prompt engineering training for translators, where students are taught how to express their disciplinary knowledge in prompts and how to critically evaluate AI-generated translations.

In assessment, Wang et al. (2025) develop an Automatic Question Generation (AQG) system that uses a prompt pattern grounded in a teacher-developed knowledge base. Teachers and researchers collaboratively code question characteristics, types, and definitions, and experienced teachers validate this knowledge base through semi-structured interviews. These characteristics then inform the design of prompts used with an LLM to generate questions from specified texts. The aim is to produce questions that align with educational standards and are comparable in quality to manually authored items.

Ching et al. focus on MCQ generation for personalised learning. Their MCQGen framework combines an LLM with RAG and a prompt engineering workflow that uses chain-of-thought prompting to generate draft questions and self-refine prompting to iteratively improve them. The system draws on a knowledge base of instructor and student-generated MCQs and is designed to support blended and flipped classroom models where frequent, targeted questioning is central.

These studies show how prompt engineering is being operationalised in concrete educational tools and raise questions about how such frameworks can be translated into teaching practice and curriculum design.

Methods

This paper employs an integrative review approach focusing on a small number of studies that exemplify different ways of formalising prompt engineering (Cronin & George, 2023). The aim is not to map the entire field but to examine how selected frameworks are constructed and what they afford educationally. The five focal papers were chosen based on the following criteria:

1. Explicit framework or pattern structure: the study introduces a catalogue, sequence, knowledge-base-driven pattern, or prompt workflow, rather than merely showing isolated prompts
2. Clear educational relevance: the framework is either used directly in an educational context (e.g. translator education, question generation for learning) or is easily transferable to higher education settings (e.g. pattern sequences for design reasoning)
3. Diversity of domains: together the studies span software engineering, software architecture, translation, and assessment/personalised learning, allowing for cross-domain synthesis

The five focal studies are White et al. (2023) (prompt pattern catalogue), Maranhão and Guerra (2024) (prompt pattern sequence), Zhang et al. (2025) (student translators' prompting behaviours), Wang et al. (2025) (AQG using a prompt pattern and teacher knowledge base), and Hang et al. (2024) (MCQGen with RAG and prompt engineering workflow). Each paper was read with attention to:

- how prompts and prompt engineering are conceptualised
- the structure and components of the proposed framework or pattern
- how the framework is applied, evaluated, or illustrated
- the educational implications explicitly mentioned or reasonably implied, especially in relation to teaching, assessment, and curriculum design

A thematic synthesis was then conducted to identify recurring forms of educational value, which informed the structure of the findings and discussion.

Results

The analysis reveals three main forms of educational value associated with the reviewed prompt engineering frameworks: (1) value for teaching, particularly in making prompt design explicit and teachable; (2) value for assessment, through supporting high-quality, scalable question generation; and (3) value for curriculum design, by foregrounding prompt engineering as a transversal competence and making AI use inspectable and assessable. Table 1 presents the analysis of the five studies.

Table 1. Analysis of the five studies

Author(s) & Year	Domain / Context	Type of Framework / Pattern	Key Components	Purpose of Framework	Educational Value
White et al. (2023)	LLM-based software development	Prompt Pattern Catalogue	• Pattern name, intent, context, motivation • Structure, examples, consequences • Categories: output formatting, reasoning, personas, error handling	To document reusable solutions for prompting LLMs and improve control, quality, and consistency of AI outputs	• Provides explicit *teaching vocabulary* for discussing prompts• Supports instruction in structuring prompts deliberately • Enables pattern-based reflective practice for students learning coding or technical writing • Suggests use of prompts as controllable, teachable design artefacts

Author(s) & Year	Domain / Context	Type of Framework / Pattern	Key Components	Purpose of Framework	Educational Value
Maranhão & Guerra (2024)	Software architecture education and industrial cases	Pattern Sequence	• Persona definition • Architectural context • Quality attribute exploration • Technical premises • Uncertain requirement statement • Extended pattern form with examples and consequences	To support complex architectural reasoning through stepwise prompting in LLM-assisted design	• Models expert reasoning and decision-making as a sequence • Scaffolds the teaching of complex, multi-step prompts • Supports reflective comparison between human and AI architectural thinking • Helps curriculum designers integrate structured prompt reasoning
Zhang et al. (2025)	Translator education / language studies	Empirical analysis of student-generated prompts	• Classification of prompt types • Analysis of prompt structure (length, detail, formality) • Examination of disciplinary concepts embedded in prompts	To understand how students naturally prompt GenAI tools and to identify gaps in prompt literacy	• Highlights deficiencies in intuitive student prompting • Identifies need for explicit prompt engineering instruction • Offers data-driven categories useful for designing prompt training modules • Informs curriculum revisions in translation and language-related fields
Wang et al. (2025)	Educational assessment / question design	Prompt Pattern, Teacher Knowledge Base	• Knowledge base of question types, characteristics, definitions • Prompt pattern aligned with validated teacher expertise • LLM-driven AQG workflow	To generate high-quality questions that align with educational standards and approximate manually authored items	• Encodes teacher expertise into prompts, preserving pedagogical quality • Supports scalable assessment design across courses • Reduces workload while maintaining validity • Acts as an instructional resource in assessment literacy courses
Hang et al. (2024)	Blended learning / flipped classroom / EdTech	MCQGen Framework	• Retrieval-Augmented Generation (RAG)	To produce contextually relevant MCQs tailored for	• Embeds prompt engineering in an end-to-end assessment workflow

Author(s) & Year	Domain / Context	Type of Framework / Pattern	Key Components	Purpose of Framework	Educational Value
			• Chain-of-thought prompting • Self-refine prompting • Knowledge base of instructor- and student-generated MCQs	personalised learning and different stages of blended learning	• Demonstrates how iterative prompting improves question quality • Provides model for integrating GenAI tools into teaching and learning activities • Useful for curriculum designers aiming to include AI-based assessment activities

Overall, the five studies show that prompt engineering is increasingly treated as a structured and purposeful practice rather than an informal or technical activity. Although the studies are situated in different educational contexts, they share a common emphasis on making prompts explicit, systematic, and aligned with teaching and learning goals. Prompts are consistently presented as design elements that can be taught, examined, and improved, rather than as trial-and-error inputs to AI systems.

The analysis also indicates that the reviewed frameworks contribute at different levels of educational practice, including classroom teaching, assessment design, and curriculum planning. To clarify these contributions, the next section discusses the findings in terms of three main forms of educational value, i.e. value for teaching, value for assessment, and value for curriculum design.

Educational Value for Teaching: Making Prompt Design Explicit

A first form of value lies in the way frameworks make prompting practices visible and discussable. White et al.'s (2023) pattern catalogue provides a structured vocabulary for describing common prompting strategies. Instead of vague advice to "be specific", educators can refer to patterns such as stepwise reasoning, persona definition, or result verification. In a classroom, this allows prompts to be treated as designed artefacts. Students can be asked which pattern they are using, why they chose it, and how they might combine it with another pattern. Because each pattern comes with an explanation of context, motivation, and consequences, the catalogue naturally lends itself to case-based teaching, where students critique example prompts and adapt patterns to new tasks.

Maranhão and Guerra's (2024) pattern sequence extends this explicitness to multi-step reasoning. Their sequence decomposes expert architectural thinking into a series of prompts that establish roles, clarify context, identify quality attributes, and surface uncertainties. For teaching, this can be repurposed as a scaffold. Students learning architecture could first work through the sequence manually (without AI), and then introduce the LLM as a dialogue partner to compare their own reasoning with AI-assisted suggestions. In this way, the sequence does not replace human judgement but externalises it and thereby makes it easier to teach.

Zhang et al.'s (2025) translation study, although not offering a new framework, highlights why such explicit scaffolds are needed. Students were found to use prompts that were short, informal, and often lacking in contextual detail, even when they held relevant theoretical knowledge. This suggests that students are not automatically able to transfer disciplinary concepts into prompt design. From an educational perspective, frameworks and patterns can bridge this gap by giving students structured ways of articulating disciplinary knowledge as prompt elements. This includes specifying target audiences, genre conventions, or quality standards rather than relying on the model to infer them.

Across these three papers, the educational value for teaching lies in the explicitation of prompting strategies. Pattern catalogues and sequences provide a shared language and exemplar structures, while empirical work reveals the distance between current student practices and the kinds of structured prompting that could better support learning.

Educational Value for Assessment: Supporting Scalable, High-quality Question Design

A second form of educational value emerges in the domain of assessment, particularly around question generation. Wang et al. (2025) start from a familiar problem, which is that constructing good questions is essential but time-intensive and demanding. Teachers must consider question types, cognitive levels, and alignment with learning outcomes, and these considerations are often only partially documented. Their AQG system addresses this challenge by combining a prompt pattern with a teacher knowledge base. The knowledge base captures practical wisdom about question characteristics (e.g. difficulty, structure, focus) and is validated with experienced teachers. This knowledge is then used to shape prompts for an LLM that generates questions from input texts.

Educationally, the system has at least three benefits. First, it demonstrates that prompt engineering can be used to encode and reuse human assessment expertise, rather than bypass it. Second, it offers teachers a way to generate questions at scale while maintaining a connection to standards they recognise as legitimate. Third, the knowledge base and prompt pattern can themselves become learning resources in assessment-related courses, where students can inspect how question characteristics are defined and how they inform prompt design.

Hang et al.'s (2024) MCQGen framework similarly situates prompt engineering at the core of assessment generation, but with a stronger emphasis on iterative refinement. Their system uses chain-of-thought prompting to have the model reason about question and distractor design, and self-refine prompting to ask the model to critique and improve its own output. By adding retrieval-augmented generation, MCQGen draws from a curated knowledge base of MCQs and course materials, thereby grounding prompts in existing content.

Here the educational value is twofold. At the system level, MCQGen provides a way to generate MCQs tailored to different phases of blended and flipped learning (pre-class, in-class, post-class), supporting personalisation and timely feedback. At the pedagogical level, the prompt engineering workflow embodies a process of critical self-evaluation, where the AI is instructed to inspect and revise its own output according to given criteria. This process can be made visible to instructors and even to students, who might be asked to critique the self-refinement steps or suggest alternative refinements.

Both systems show that prompt engineering frameworks can be used not only to generate more questions, but to support better conversations about what counts as a good question and how assessment design decisions are embedded in prompt logic.

Educational Value for Curriculum Design: Embedding Prompt Engineering as a Transversal Competence

A third form of educational value lies at the programme and curriculum level. Across the five focal papers, prompt engineering appears in quite different guises. It is presented as a pattern catalogue for software development, a sequence for architecture decision-making, a focus of professional training in translation, and a core component of systems for question generation and personalised learning. This diversity suggests that prompt engineering is not restricted to computer science or data science. It is becoming a transversal competence that can be adapted to the epistemic and practical demands of different disciplines.

For curriculum designers, these frameworks provide concrete anchors for designing AI-related learning outcomes. A generic first-year module might introduce fundamental prompt patterns. These include defining context, specifying output formats, and asking for stepwise reasoning. Such a module can draw loosely on catalogues such as that proposed by White et al. (2023). More advanced, discipline-specific modules could then introduce domain-tailored frameworks: for example, architecture courses could teach the use of persona and context sequences adapted from Maranhão and Guerra (2024), while translation courses might develop training activities grounded in Zhang et al.'s (2025) classification of student prompts.

The frameworks also support assessment of AI use. Because patterns and sequences make prompt logic explicit, they allow educators to ask students to submit prompts alongside their work and to evaluate them using clear criteria. These criteria include whether the prompt is aligned with the task, whether it adequately expresses disciplinary concepts, and whether it reflects critical awareness of AI limitations. This approach can help shift institutional debates from whether students use GenAI to how well they design and reflect on their use.

Finally, the reviewed frameworks can inform staff development and policy. They show examples of AI being used as a collaborator in design and assessment tasks rather than as a replacement for human expertise. This framing can support institutional efforts to articulate principles for responsible GenAI use, emphasising co-design and transparency rather than automation alone.

Discussion

The findings suggest that prompt engineering frameworks have significant potential to contribute to higher education in three interconnected areas, which are teaching, assessment, and curriculum design. From an educational theory perspective, these contributions relate to how learners construct understanding, regulate their thinking, and engage meaningfully with tools rather than use them uncritically. At the same time, they raise questions about what kinds of knowledge and dispositions educators want students to develop in relation to GenAI.

From a teaching perspective, prompt patterns and sequences help to demystify GenAI interaction. When students see prompts as programmable structures rather than opaque "magic spells", they can begin to reason about them. They can consider why a particular persona is chosen, which contextual details are necessary, and how a sequence of prompts guides the model through complex tasks. This supports a constructivist view of learning, where understanding is actively built through reasoning and refinement rather than passive reception of AI outputs (Grubaugh et al., 2023). This positions prompt engineering as an object of metacognitive reflection, aligning with broader educational goals such as critical thinking and self-regulation.

From an assessment perspective, frameworks like those of Wang et al. (2025) and Hang et al. (2024) show that prompt engineering can help align AI-generated assessment items with human-defined standards. This is particularly important in an era where purely automated, black-box question generation could easily undermine trust in assessment processes. By making the prompt logic explicit, educators can evaluate whether the system is reinforcing or eroding desirable conceptions of quality questions and fair assessment.

At the curriculum level, recognising prompt engineering as a transversal competence suggests that GenAI should not be confined to isolated "AI modules". This reflects curriculum design principles that stress coherence, progression, and disciplinary relevance across programmes. Instead, different programmes can embed prompt-related learning outcomes in ways that reflect their disciplinary practices. In translator education, this might mean learning to express genre, register, and pragmatic intentions in prompts. In architecture, it involves articulating technical constraints and quality attributes. In educational technology, it requires designing and critiquing AI-supported assessments.

There are, however, tensions to navigate. If prompt engineering is framed only as a way to maximise efficiency or "get the right answer" from AI, educational efforts may inadvertently encourage instrumental use and reliance. Educational theories of deep and surface learning caution against such performance-oriented approaches that prioritise output over understanding. The reviewed work also points to another direction i.e. using prompt frameworks to slow down interactions, to encourage explanation, self-critique, and alignment with human values. MCQGen's self-refine workflow is a good example of this slower, reflexive approach, as is the use of pattern sequences to structure design reasoning rather than to shortcut it.

The educational value of prompt engineering frameworks therefore hinges not just on their technical sophistication, but on how they are introduced and used. When grounded in sound educational principles such as learner agency, transparency, and reflective practice, these frameworks can support meaningful integration of GenAI into higher education. If they are taught as tools for enhancing agency, transparency, and critical engagement, they can support meaningful integration of GenAI into higher education. If they are taught primarily as hacks for generating more content more quickly, they risk reinforcing superficial engagement and dependency on AI.

Conclusion

This paper has reviewed five prompt engineering frameworks and prompt pattern–based approaches with clear relevance to higher education, focusing on the question of what educational value they demonstrate or imply. Across domains as varied as software development, architecture, translation, and assessment design, the analysis suggests that prompt engineering frameworks can play a significant role in helping higher education to engage constructively with GenAI.

For teaching, pattern catalogues and sequences provide structures that make prompting strategies explicit, discussable, and teachable. They offer students a way to see prompts as designed entities rather than opaque tricks, and they enable educators to externalise expert reasoning for critique and adaptation. For assessment, frameworks embedded in AQG and MCQ generation systems show that prompt engineering can support scalable, high-quality question design while keeping human-defined standards at the centre. For curriculum design, the diversity of applications signals that prompt engineering is emerging as a transversal competence, one that can be incorporated into AI literacy initiatives and discipline-specific modules alike.

At the same time, evidence from current student practice reminds us that without guidance, learners are likely to rely on intuitive prompting that may underuse their disciplinary knowledge and underplay critical reflection. The reviewed frameworks provide one route towards more intentional practice, but they must be embedded in thoughtful pedagogy that emphasises reflection, ethics, and human judgement.

Ultimately, the educational value of prompt engineering frameworks is not merely that they help obtain better answers from AI. Their deeper value lies in how they can help universities cultivate graduates who know how to design their interactions with AI. These graduates are able to structure prompts that reflect disciplinary understanding, interrogate AI outputs, and situate GenAI as a tool within broader human purposes. If higher education chooses to treat prompt engineering as a serious object of teaching, assessment, and curriculum design, the frameworks reviewed here offer a set of starting points for that work.

Limitations and Future Research

This review has several limitations that should be acknowledged. First, it focuses on five studies selected for their explicit framework contributions. While this allows for depth of analysis, it does not capture the full range of work on prompt engineering in education or related domains. Other frameworks, including those emerging from different disciplines or languages, may bring additional insights or challenge the patterns identified here.

Second, two of the major framework papers are rooted in software engineering and architecture. Their translation to other fields such as social sciences, humanities, or health professions is conceptually plausible but empirically untested. Similarly, while the AQG and MCQGen systems are clearly educational, their use cases and evaluation contexts may not generalise to institutions with different assessment cultures or technological infrastructures.

Third, evidence regarding learning outcomes from teaching prompt engineering is still scarce. Zhang et al. (2025) provide a valuable picture of current student practices and argue for specialised training, but do not yet evaluate the impact of such training. Likewise, studies on AQG and MCQGen focus on item quality and system performance rather than on how students' understanding or skills evolve when they engage with these systems.

These limitations point to several avenues for future research:

1. Design and evaluation of prompt engineering instruction in real higher education courses, using pattern catalogues and sequences as teaching tools, with systematic measurement of changes in student prompting behaviour, critical thinking, and task performance.
2. Development of assessment instruments for prompt engineering competence, potentially drawing on the structures used in the reviewed frameworks to define criteria such as contextualisation, structure, reflective refinement, and ethical awareness.
3. Discipline-specific adaptations and comparative studies to explore how frameworks need to be modified for different epistemic cultures, and whether certain patterns are more or less effective in particular fields.

4. Equity-focused research examining how prompt engineering frameworks interact with differences in students' language proficiency, digital literacy, and access to AI tools. Explicit frameworks may reduce uncertainty and support less confident students, but they may also introduce new barriers if presented in overly technical ways.
5. Institutional and ethical investigations into how pattern-based prompt engineering can be integrated into academic integrity policies, accreditation processes, and professional standards, particularly in disciplines where accountability and transparency are critical.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

We would like to thank Open University Malaysia for their support throughout the research process.

References

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, *15*(3), 1-45. https://doi.org/10.1145/3641289

Coplien, J. O. (1998). *A generative development process pattern language* (pp. 243-300). Cambridge University Press, New York.

Cronin, M. A., & George, E. (2023). The why and how of the integrative review. *Organizational Research Methods*, *26*(1), 168-192. https://doi.org/10.1177/1094428120935507

Geroimenko, V. (2025). Key Principles of Good Prompt Design. In *The Essential Guide to Prompt Engineering: Key Principles, Techniques, Challenges, and Security Risks* (pp. 17-36). Cham: Springer Nature Switzerland.

Grubaugh, S., Levitt, G., & Deever, D. (2023). Harnessing AI to power constructivist learning: An evolution in educational methodologies. *EIKI Journal of Effective Teaching Methods*, *1*(3). 81-83. https://doi.org/10.59652/jetm.v1i3.43

Hang, C. N., Tan, C. W., & Yu, P. D. (2024). MCQGen: A large language model-driven MCQ generator for personalized learning. *IEEE Access*, *12*, 102261-102273. https://doi.org/10.1109/ACCESS.2024.3420709

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*(1), 7. https://doi.org/10.1186/s41239-025-00503-7

Maranhão, J. J., & Guerra, E. M. (2024, July). A prompt pattern sequence approach to apply generative AI in assisting software architecture decision-making. In *Proceedings of the 29th European Conference on Pattern Languages of Programs, People, and Practices* (pp. 1-12). https://doi.org/10.1145/3698322.3698324

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2), 1-24. https://doi.org/10.53761/1.20.02.07

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). The prompt report: a systematic survey of prompt engineering techniques. *arXiv preprint arXiv:2406.06608*.

Wang, L., Song, R., Guo, W., & Yang, H. (2025). Exploring prompt pattern for generative artificial intelligence in automatic question generation. *Interactive Learning Environments*, *33*(3), 2559-2584. https://doi.org/10.1080/10494820.2024.2412082

White, J. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*.

Xu, Z., Peng, K., Ding, L., Tao, D., & Lu, X. (2024). Take care of your prompt bias! investigating and mitigating prompt bias in factual knowledge extraction. *arXiv preprint arXiv:2403.09963*.

Prompt Engineering Frameworks and Their Educational Value in Higher Education. *Innovative Teaching and Learning Journal*, *9*(2), 495–505.

Zhang, J., Zhao, X., & Doherty, S. (2025, June). Prompt engineering in translation: How do student translators leverage GenAI tools for translation tasks. In *Proceedings of Machine Translation Summit XX: Volume 1* (pp. 420-431).