Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140625

Statistical Prompt Engineering Framework for Multiple-Choice Questions for Film Critique Pedagogy

Sahar R. Hamzah

Department of Education and Curricula, Faculty of Digital Education, Arabian Gulf University, Building 293, Road 2904, Block 329, Manama, Kingdom of Bahrain

Received: 2 Sep. 2025, Revised: 12 Oct. 2025, Accepted: 22 Oct. 2025.
Published online: 1 Nov. 2025.

Abstract: This research presents a statistical prompt engineering framework for developing film and literary critique. The framework is designed to generate multiple-choice questions that scaffold students through the cognitive skills needed for film critique. Film criticism is a major component of film and media studies. Students are often required to learn and improve their critique skills through performing critique without first having internalized the underlying cognitive skills required to perform critiques. The purpose of this study is to develop a replicable prompt engineering framework that demonstrates how instructors can use generative artificial intelligence to create multiple-choice questions that will support the step-by-step development of critique skills. Using a design-based methodology, the study breaks down film criticism into teachable cognitive components and aligns these components with Bloom's Taxonomy. A master prompt was developed to generate a 50-question MCQ. The questions were of mixed difficulty and designed to support multiple interpretations and encourage comparative judgment. The master prompt was created using an iterative meta-prompting process to refine the structure, the constraints, and output strategies. The contribution of the study is a theoretically grounded master prompt and step-by-step instructor workflow that expands higher-order multiple-choice question design into humanities contexts, where such approaches are underexplored. The research also presents a prompt engineering framework for item and frequency analysis of the multiple-choice questions. The study concludes that, when constrained by pedagogical intent and explicit design rules, generative artificial intelligence can be used as an instructional design assistant that encourages the development of critical thinking in humanities education.

Keywords: Film critique, Pedagogy, Multiple-choice questions, Prompt engineering, Bloom's Taxonomy, Generative artificial intelligence, Scaffolding, Design-based research.

1. Introduction

Film criticism is a major component in film and media studies, functioning as an analytical practice through which students develop visual literacy, interpretive judgment, and critical reasoning. Traditionally, film pedagogy has placed its primary emphasis on essay writing, seminar discussion, and audiovisual production as the primary means through which students learn to analyze and evaluate cinematic meaning, context, and form [1, 2]. These approaches view criticism as a cognitive activity that requires evaluation, synthesis, and the ability to justify interpretive claims through formal and contextual evidence. Structured instructional support is often essential in helping students acquire these complex disciplinary skills, especially when those skills involve judgment, ambiguity, and evaluative reasoning [3-6].

In education in general, multiple-choice questions (MCQs) are one of the most widely used assessment formats due to their efficiency, scalability, and capacity to cover a broad range of content [6, 7]. MCQs are often criticized for eliciting recall rather than critical thinking. However, MCQs can be designed to require critical thinking, including analysis, evaluation, and decision-making. This is especially seen when aligned with Bloom's Taxonomy. This is also seen when the MCQs are constructed to require choosing between plausible alternatives [5, 7]. MCQs can support learning when they are used in a formative manner whilst embedded within instructional scaffolding and designed around disciplinary reasoning rather than to test knowledge as part of summative evaluations [6].

In spite of recent research developments in this field, there is a gap between assessment research on MCQs and pedagogical research on film criticism; the key element of this paper. Literature on film pedagogy largely focuses on essays and audiovisual criticism as a means for developing students' critical skills and thinking methodology. There are very few examples of well-structured, low-stakes formative learning activities that support critique development [1, 2]. As a result, students are often expected to execute complex acts of film criticism without sufficient opportunities to demonstrate a strong foundation in critical areas of underlying cognitive work, such as:

*Corresponding author e-mail: saharrh@agu.edu.bh

- Identifying formal cues
- Comparing interpretive possibilities
- Evaluating the strength of competing readings in a guided and incremental manner.

This gap is of pedagogical significance for several reasons. First, meaning in film criticism relies on contingency and interpretation and is dependent on content. This makes it well-suited to scaffolded learning approaches based on constructivist theory and instructional scaffolding [8, 9]. Second, cognitive complexity and difficulty are separable. As such, MCQs can be purposefully designed to target critical thinking [6]. The strategic use of Generative Artificial Intelligence (GAI) applications makes it easier to implement within instructional planning to connect learning outcomes, cognitive objectives, and assessment design in transparent and repeatable ways.

This study introduces an MCQ prompt framework designed to be used to create scaffolded activities to support the development of film critique skills. The framework is informed by Bloom's Taxonomy, which is a hierarchical framework that categorizes learning and cognitive thinking, as well as by scaffolding theory, and research on metacognitive thinking [3, 4]. This framework is designed to aid instructors in generating interpretive MCQs that are aligned with specific cognitive and aesthetic critique skills. This research aims to contribute to film pedagogy by demonstrating how low-stakes assessment tools such as MCQs can support the development of critical film analysis skills and reduce the barriers of entry into the practice of film criticism.

1.1 Film Criticism Pedagogy

Research suggests that film criticism taught through pedagogical activities such as writing essays, discussions, peer evaluation, and film commentary can often be experienced by students, especially those new to critique, as high-stakes learning environments [1, 2]. Class discussion and peer critique are often framed as being dialogic and exploratory as they require students to publicly articulate their interpretive claims, justify their judgments, and negotiate disagreements in real time. For novice critics, this places the performance of critique itself under evaluation. While students are attempting to interpret a film, they are also being judged on the sophistication, confidence, and legitimacy of their critical voice. As a result, critique becomes a source of anxiety rather than inquiry. Rather than engaging in formal analysis or interpretive reasoning, students may instead choose safer positions by stating personal preferences such as "I liked it" or "I did not like it" or simply listing perceived flaws in the films.

Pedagogical research in film and art criticism emphasizes that critique is a learned practice, and not an innate skill. It is developed through sustained training of both perception and cognition. Criticism is a developmental process. Learners are formally taught how to elevate their initial response into a more disciplined judgment. This teaches them to learn the difference between subjective reaction and reasoned evaluation in their analytical approach [10]. Critique is a skill that requires practice and discipline to develop in a reasoned way by analyzing form, articulating meaning, and weighing up evidence [1]. Even when rubrics are provided to them, students must still train the eye to recognize the cinematic cues given to them in a consistent manner. They must then interpret these cues in terms of significance and justify evaluative claims through evidence rather than their own opinion. Broadly speaking, this training is expected to develop through time and practice, without articulating what skills are needed in order for students to train their eye.

When considering how to introduce low-stakes opportunities for students to rehearse their critique, the literature offers few established models that connect the cognitive skills of film criticism to specific instructional tools. Although film pedagogy has theorized the nature of critical viewing and interpretation, there are no frameworks that map critique skills onto scalable cognitive practice activities. Though rubrics are used to help identify areas of attention for critique, critique is not only a matter of technical skill. It is also a matter of attitudinal formation. Teaching critique requires guiding students in the acquisition of the perceptual and analytical skills in addition to cultivating dispositions such as comfort with ambiguity, openness to multiple interpretations, and an understanding of objectivity as reasoned judgment instead of neutrality. In this way, students who are new to critique are less likely to assume that criticism involves expressing personal preference or identifying what is “wrong” with a text.

Research does recognize the existence of an established instructional tool capable of supporting this type of cognitive rehearsal...the multiple-choice question [5-7, 10]. Assessment literature demonstrates that the format of MCQs are cognitively flexible and can be engineered to engage with complex reasoning, interpretive judgment, and evaluative decision-making [12, 13]. When designed with an intention to engage specific cognitive skills, MCQs can act as structured prompts that guide learners through disciplinary ways of thinking.

1.2 MCQs and Cognitive Thinking

MCQ research in medical, nursing, and health sciences education, has extensively examined how MCQs can be designed to

assess and support higher-order cognitive skills [5-7, 11, 12]. If MCQs are designed using Bloom's Taxonomy, a system can be incorporated to create questions aimed at specific cognitive orders. Bloom's Taxonomy organizes learning objectives into six progressively complex cognitive levels. These levels consist of remembering, understanding, applying, analyzing, evaluating, and creating. Lower-order cognitive skills involve remembering and understanding information, and higher-order cognitive skills involve applying, analyzing, evaluating, and creating based on that information. This structure helps educators design material that promotes deeper levels of thinking and intellectual engagement.

Research supports the pedagogical value of deliberately designing assessments to target higher-order cognitive processes. Jensen et al. demonstrate that students who were assessed throughout a semester using higher-level questions aligned with application, analysis, and evaluation not only performed better on high-level exam items, but also outperformed their peers on lower-level recall items in a cumulative final examination [11]. This suggests that sustained exposure to higher-order assessment prompts will encourage deeper conceptual processing that supports both higher and lower cognitive levels. This lends support to the logic hierarchy of Bloom's Taxonomy. Concurrently, assessment research demonstrates that students have a greater tendency to score better on lower-order questions than on higher-order questions [11]. This pattern justifies the inclusion of lower-order questions as a mechanism of confidence-building entry points within mixed-difficulty assessments.

When assigning a correct Bloom level to questions, faculty find that there is usually some variation amongst themselves in their designated levels. This is reflective of the interpretive nature of cognitive-level classification [12]. However, this variability does not invalidate the usefulness of Bloom's Taxonomy for MCQs. It demonstrates how the taxonomy should function as an instructional planning tool. It should not be used as a rigid labeling system. When instructors design MCQs with varied cognitive targets, students are still engaged with a spectrum of cognitive demands. Using Bloom levels just ensures that there is cognitive diversity within the assessment. However, in order for this to be achieved, the question prompt (stem) and the answer choices (distractors) must be carefully constructed [7, 13].

Research shows that what makes a question higher-order is not whether the question is multiple-choice, essay, or short answer, but the kind of thinking students must use to choose between the answer options [14, 15]. Haladyna provides recognized and established guidelines for item-writing, referring to the writing process of MCQs. These guidelines suggest that the quality of the MCQs depends on adhering to these validated item-writing principles, instead of the format itself. These guidelines include that each question should present a clear and focused problem in the stem and should avoid unnecessary irrelevant or complex information. The question should also be free of grammatical or logical errors. There should not be any length-based cues that might reveal the correct answer. [14-16] The answer options should have a parallel structure across options. The answer options should also align directly with the specified learning objectives. These guidelines were created to ensure the items are able to engage the intended cognitive levels.

The incorrect distractor answer options should be plausible and meaningful alternative options. These options could reflect common misconceptions, partial understanding, or competing interpretations. They should not be trivially incorrect or too logically similar to other options. They should not be structured it a way that makes it stand out, such as being much shorter or longer than other answers. When distractors are poorly constructed in this way, students may select the right option through an element of pattern recognition instead of reasoning. Studies in nursing and medical education demonstrate that flawed items artificially distort the perceived difficulty of the MCQ and therefore compromise its validity [7, 16]. Students must be able to discriminate between competing lines of reasoning in order to fully engage in critical thinking tasks. MCQs being designed with this level of detail, and adherence to scholarly considerations of higher- and lower-order thinking challenges the notion that MCQs only address recall.

Research also suggests that a question can require complex thinking without being more difficult. A difficult question is not automatically a higher-order one [6]. MCQs can be designed to push analytical or evaluative thinking without increasing stakes or penalizing learners who are still developing their skills. This body of research however is largely discipline-bound. The majority of research that models how to design MCQs that test higher-order thinking comes from medical and health science fields. Very little of this research has been done in humanities fields such as film studies, even though they rely heavily on interpretation and judgment [17-19].

Humanities education tends to underutilize MCQs and often dismiss them as being incompatible with interpretive disciplines. Bloom's Taxonomy is frequently cited as a theoretical justification for higher-order MCQs for evaluative thinking. Yet, there is little published guidance on how this can be put into practice for disciplines that are characterized by the subjectivity of aesthetic judgement or multiple interpretations.

Research in the discipline does not adequately explain how MCQs could be designed to teach students how to interpret, compare different readings, or make evaluative judgments without forcing one single "correct" answer. Because of this, humanities teachers do not have proven guidelines for using MCQs as learning tools in this way. Students are then expected to produce critiques without first practicing the cognitive micro-decisions that critique requires. When there is no structured

practice between introduction and advanced critique, students cannot easily reach deep levels of understanding because they are spending most of their cognitive effort just trying to survive the task rather than exploring ideas. Well-designed MCQs can act like non-performative practices that let students practice their higher-order thinking in a low-stakes format.

1.3 GAI and the MCQ

GAI has been shown to be an efficient tool in higher education to aid instructors in creating MCQs [18-20]. Studies in medical and health education suggest that GAI can generate high quality MCQs that align with cognitive taxonomies and learning outcomes when the tool is given carefully structured prompts and the outputs are reviewed and refined [18, 19]. In these studies, GAI-produced MCQs are not autonomous assessors of students. Studies emphasize that the pedagogical value of the GAI content depends greatly on human oversight. This type of use of GAI assessment frameworks have been explored almost exclusively within medicinal and health disciplines, leaving the use of GAI for MCQ design in the humanities especially unexamined.

The focus of GAI within the humanities has largely been shaped by debates surrounding authorship, originality, creativity, and the epistemic value of AI-assisted work. Studies focus on interrogating to what degree these GAI outputs constitute creative labor, how authorship should be attributed, or how a student's academic integrity is impacted when they use GAI to produce creative works. Existing discussions tend to treat AI as a threat to authentic learning or as a shortcut that undermines critical thinking. These concerns have been particularly pronounced in disciplines such as literature, creative writing, film, and art studies, as points of interpretive authority and originality are often central to the disciplinary identity itself. Scholarship acknowledges that film criticism has undergone digital transformation with explorations of how digital platforms have affected the relationship between the critic and the audience. The digital ecosystem of film criticism has become interactive and participatory through global media forms such as blogs and social media. This has intensified debates about professionalism and authority with discussions of what makes a critic when anyone can publish their critique online [21].

Film criticism is a mature and highly concentrated field that has been characterized by dominant intellectual hubs. These changes in critique formats, with increased use of GAI and other digital approaches, indicate a need for greater collaboration with digital humanities [22]. There is extensive literature on GAI's role in producing or evaluating critique, as well as how to ascertain the artistic contribution of the critic from the use of GAI [21-24]. This reflects a gatekeeping logic in how the disciplines have viewed authority in terms of originality and authorship. Traditionally, disciplines seemingly prioritized expressive mastery in written or verbal critique over the cognitive processes through which the critical judgment itself is created. Similar debates emerge in literature studies. Scholars acknowledge the usefulness of the use of GAI in critical analysis for brainstorming ideas or identifying themes and patterns. However, skepticism still remains when debating its capacity for the nuanced interpretation needed in the discipline [23].

While these debates are necessary, they have also resulted in a narrow framing of AI primarily as a creation tool with little attention to the implications of its usage by instructors as a pedagogical instrument. As a consequence, relatively little empirical or theoretical work has examined how GAI might be used to support the instructional design, assessment creation, or structured scaffolding of disciplinary thinking in the humanities. This contrasts with medical and health sciences education where GAI has been explored as a means of reducing instructional workload while maintaining or improving assessment quality [18, 19]. GAI can help instructors by reducing cognitive load of lower-order tasks, such as routine question writing, so that they can engage in more higher-order cognitive work, such as analyzing and developing the MCQ-generated outputs to ensure that the outputs fit their teaching goals.

Interestingly, although research on MCQs and Bloom's Taxonomy within film pedagogy is limited, scholarship from other disciplines have already mapped analyzing films to cognitive taxonomies and other learning outcomes. An action-research study in a British secondary English course aligned learning outcomes from watching films with Bloom's Taxonomy [25]. This was used to explain how students may progress from recalling information about films to understanding meaning, applying concepts, and evaluating through comparison or critique, to ultimately create new interpretations. These interpretations connect with learning outcomes within the English course.

Similarly, research establishes film as an effective teaching tool to encourage analysis, synthesis, and evaluation. This is particularly seen when students engage in tasks such as group discussions, oral presentations, writing reviews, and critiques connected to film content. These studies suggest that active viewing of film, supported by structured cognitive tasks, supports developing critical thinking and higher-order cognition [25]. Language and education fields already design assignments that intentionally map what students should extract from films onto hierarchical cognitive skill models. However, even in these fields, such mapping has almost exclusively been achieved through open-ended tasks such as essays, projects, or discussions, not through structured formats like MCQs. This shows that film studies and Bloom's Taxonomy work well together in principle, but there are no well-established, tested methods for using MCQs to teach or assess this kind of thinking in film studies, or how instructors might intentionally harness GAI to support the development of interpretive skills. Additionally, again they rely on high-stakes assessment formats such as essays, critiques, and presentations that presuppose students'

fluency in disciplinary reasoning without offering effective low-stakes opportunities for practice [26].

1.4 Prompt Engineering as Pedagogical Design

Recent scholarship on GAI explores prompt engineering as a place where the instructor's pedagogical intent, cognitive scaffolding, and creative production intersect. This perspective treats prompting as a type of pedagogical skill, as effective prompts encode learning objectives, constraints, and expectations in ways that impact the quality of generated outputs but also the cognitive processes required to engage with them [27]. Writing prompts is a teaching activity in itself, since when instructors design prompts, they are turning learning goals into specific thinking tasks. This can be a creative part of teaching. Instructors must experiment, adjust, and refine prompts based on the skills they want students to practice. This creativity is similar to how instructors design any non-GAI-based learning material. When instructors design prompts that produce thoughtful MCQs, they can be creatively involved in how students practice critical thinking within their discipline.

Prompt engineering can be used as a form of constraint-based creativity. This concept discusses creativity as not emerging from being given unlimited freedom. It comes from working within carefully designed constraints [28]. Within assignments, the instructions are the constraints. Instructors have the ability to design these constraints to influence how students construct meaning in their discipline. The type of creativity required to design effective assignments is the same type of creativity required to design strong prompts [29]. For film criticism, constraints can narrow the interpretive attention towards specific elements such as framing, cinematography, sound, performance, or narrative. Without adding these types of constraints when prompt engineering learning material, GAI tools are more likely to generate surface-level recognition tasks or reduce multiple interpretations into simple right or wrong answers. Prompt engineered MCQs can function as such constraints for both instructors and students. Instructors creatively design the conditions under which questions are generated, and students engage creatively by navigating structured interpretive possibilities. Well-constructed MCQs can support creative and critical thinking by delimiting the space of possible interpretations while still preserving multiple interpretations and allowing creativity to take the form of reasoned selection, comparison, and justification. When exploring question asking and prompt engineering, it is also suggested that the higher the cognitive order, the higher the creativity [29]. However, literature offers little guidance on how instructors can operationalize GAI skills such as rapid content generation and variation to acquire the implied benefits and create repeatable, pedagogical assessment practices for interpretive disciplines.

By specifying in the prompts constraints on the cognitive thinking target and the disciplinary lens, instructors can iteratively generate, test, and refine MCQs that enact the kinds of thinking they seek to develop. This approach ensures that the pedagogical judgment needed is not outsourced to the GAI. It is the instructors who design the GAI prompts to shape the conditions under which student cognition occurs. With this approach, prompt engineering can support MCQs as an interpretive system. By providing multiple defensible interpretations, formal analytic vocabulary, and contextual cues within answer choices, MCQs can model the comparative reasoning and evaluative discrimination that is necessary for film criticism. The constraints of the MCQs can be designed to require students to select between multiple justifiable answers. This mirrors the act of critiquing. Critics weigh competing readings and justify their judgments based on the evidence. This design approach is constructivist in nature, and allows for judgments to be scaffolded as it allows learners to internalize critical practices and construct meaning. This repeated guided engagement is able to maintain low-stakes conditions that encourage exploratory learning and interpretive risk-taking.

Assessment research demonstrates that MCQs can support higher-order thinking, but this work remains largely confined to domains with determinate reasoning structures. Meanwhile, discussions of prompt engineering highlight the pedagogical potential of constraint-based design but rarely connect it to pedagogy for humanities assessment or critique. This research addresses this gap by integrating film pedagogy, assessment theory, and prompt engineering into a unified framework for designing MCQs as scaffolded learning activities for film criticism. Instead of evaluation of GAI based on authorship, it is necessary to assess how GAI can help students learn. This research examines how GAI can help students practice disciplinary thinking in low-stakes, scaffolding learning environments before or alongside more performative, high-stakes methods such as in class critiques or essay.

2. Materials and Methods

This study uses a design-based research methodology centered on the creation and analysis of a pedagogical ChatGPT master prompt. The research examines how pedagogical design choices can operationalize theory. It explores how instructors can incorporate cognitive engagement and scaffold disciplinary thinking. The study conceptualizes the MCQ as a pedagogical intervention and design framework intended to address recurring instructional challenges in film criticism. The intervention presents a ChatGPT prompt framework for the design of low-stakes, repeatable, scaffolded, and interpretively rich MCQs. This methodology is informed by constructivist learning theory, Vygotskian scaffolding, and Bloom's Taxonomy. The MCQs are designed to be interpretive environments that allow learners to rehearse the discrete cognitive operations of film criticism.

2.1 Target Audience and Pedagogical Scope

The learning tool is instructors teaching students that are studying film or other adjacent humanities disciplines. The MCQs are aimed at high school or undergraduate university students. The framework is aimed at learners who are new to formal critique and may experience difficulty transitioning from personal opinion or fault-finding attitudes towards disciplinary analytical judgment. However, the tool is adaptable across different educational levels and contexts as instructors can tailor the questions according to their topic and theoretical focus of choice.

2.2 Case Study Selection: Pride and Prejudice

The case study used for developing the prompt was *Pride and Prejudice*. The study draws on Jane Austen's 1813 novel and two widely recognized screen adaptations, the 1995 BBC miniseries directed by Simon Langton and the 2005 feature film directed by Joe Wright. This text was chosen for several reasons. Its canonical status makes it suitable for both secondary and tertiary education contexts. The narrative exists within the curricula of a variety of disciplines. The public familiarity with the story and plot reduces extraneous cognitive load associated with plot comprehension. Existing adaptations of the work present contrasting stylistic and interpretive approaches, offering many opportunities for comparative analysis from a film production, media studies, or literary perspective. The texts and film adaptations support multiple defensible interpretations. The familiarity of the source material allows the pedagogical design to focus explicitly on how critique is performed, rather than on whether learners understand the narrative content.

2.3 Deconstructing Film Criticism into Teachable Cognitive Skills

Bloom's Taxonomy was used as a tool for building the *Pride and Prejudice* MCQ question bank. It was not used to label questions after they were written or as a checklist for categorizing them afterwards. It was used to guide the prompts in how the questions and answers were generated, as well as the cognitive demand required within the output. This process ensured that the MCQs were able to be prompted to work as progressive scaffolds that allow learners to move from perceptual noticing towards evaluative judgment within this structured environment.

Film criticism was first deconstructed into components that were then mapped to cognitive skills and Bloom levels. This allowed the MCQs to be able to be prompted to function as scaffolded learning tools. The mapping to Bloom's Taxonomy functioned to help learners move from lower-order to higher-order thinking. The GAI system is then able to generate questions and answers that scaffold a specific type of reasoning. The film critique components incorporated disciplinary language. The aim was to create questions with multiple interpretations within the question stems and answer choices. The purpose of this was to create MCQs that were able to provide structured support for students to practice analytical and evaluative techniques they may or may not already have [8]. The aim of the research was to design a framework for creating a rehearsal space that would support learners in incrementally acquiring the perceptual, analytical, and evaluative skills required of film critique.

Drawing on film pedagogy and art criticism, the following skill model in Table 1 was developed to make the cognitive operations of film critique pedagogically explicit and scaffoldable through MCQ design:

Table 1: Framework aligning film critique skill model, key criteria and Bloom level

Skill Type	Criteria	Bloom Level
Perceptual	Recall facts such as plot elements and character details in the adaptations and text. Notice mise-en-scène, including framing, sound, tone, rhythm, and performance, corresponding to Bloom's remembering and understanding levels, and forming the perceptual foundation upon which higher-order analysis depends [3, 4, 30].	Remember Understand
Analytical	Breaking scenes into formal and stylistic elements. Match production technique to output on screen, corresponding to Bloom's apply and analysis level and reflecting the structured breakdown of cinematic form central to formal film criticism [3, 4, 30, 31].	Apply Analyze
Interpretive	Reasoning and connecting cinematic form to meaning and theme, engaging analysis and evaluation by requiring learners to infer significance from formal evidence rather than rely on personal response [3, 4, 10].	Analyze Evaluate
Comparative Judgment	Weighing alternative readings or adaptation choices, reflecting evaluative cognition in which multiple plausible interpretations are considered within defined criteria [2-4].	Evaluate
Evaluative Justification	Justifying why one interpretation is more strongly supported by evidence, corresponding directly to Bloom's evaluation level and emphasizing reasoned judgment over subjective preference [3, 4].	Evaluate

Contextual	Integration of historical, cultural, and ideological awareness, requiring synthesis across textual and contextual domains. This can be associated with higher or levels of Bloom's taxonomy [1, 3, 4].	Understand Apply Analyze Evaluate
Metacognitive Awareness	Recognizing one's own interpretive stance and assumptions, or constructing an new interpretation, consistent with constructivist learning theory and reflective practice, and essential for regulating one's movement through scaffolded increasingly critical tasks [8, 9].	Create Evaluate

2.4 Higher-Order Interpretive MCQs

The higher-order interpretive MCQs were developed to have multiple defensible answers and scaffold disciplinary thought through answer choices. Questions were framed to ask which interpretation was most strongly supported or which is the best answer instead of the correct answer. Accompanying distractors could be justified in varying degrees. The purpose of this was to shift the evaluative task from correctness to reasoning quality. Answers were designed to model disciplinary film vocabulary. Answers were also designed to present common novice points of discussion or misconceptions. This higher-order MCQ format is what creates the critique rehearsal space. This aims to reflect the epistemic nature of film and literary criticism. In this way, the answer choices themselves function as instructional scaffolds that cognitively embed critique apprenticeship throughout the MCQs.

2.5 Prompt Engineering as Design Method

ChatGPT was selected as the design tool of choice due to its accessibility and its ability to support the generation and iterative refinement of the MCQs for the project. All of the outputs subject to human curation by the researcher, including revisions and ensuring the theoretical alignment. Prompt engineering was used to creating a prompt consisting of four layers:

- The cognitive target, e.g., evaluation-level reasoning, Bloom's Taxonomy.
- The disciplinary lens, e.g., mise-en-scène, cinematography, tone, adaptation choice.
- The interpretive plurality, e.g., at least two defensible readings.
- The scaffolding constraints, e.g., disciplinary language and modelled reasoning.

An example prompt used in development was:

- Generate a multiple-choice question for film criticism that assesses evaluative reasoning, includes at least two defensible interpretations, and requires selection based on cinematic evidence.

Some questions were designed to target higher-order interpretive reasoning while others were designed to be lower-order. These lower-order questions would look at recall elements such as plot details. The choice to include this type of question was to assist in building the confidence of learners and helping to reduce their cognitive load. This is due to the fact that sustained engagement with complex tasks has been found to be more effective when interspersed with lower-order items [4, 26, 32].

The questions that were more cognitively demanding were the questions designed to be the interpretive rehearsals. One example of this involved a question that was written and entered into ChatGPT to generate answer options that engage with Bloom's Analyze and Evaluate levels. The question addressed a moment of narrative inference rather than textual fact.

- Question:
  - Who informed Lady Catherine de Bourgh about the potential relationship between Elizabeth Bennet and Mr. Darcy?

At a literal level, the novel or adaptations do not specify who provides this information. The aim of the question was that with the absence of a definitive answer, an opportunity would be created for deductive reasoning requiring character analysis and an exploration of character dynamics within the text.

- Answers:
  - Mr. Darcy
  - Charlotte Lucas
  - Mr. Collins
  - Mrs. Bennet

This question was designed to prompt students to consider patterns of gossip, social behavior, and character motivation within the story. Students could, through their own interpretive judgement, make a reasonable deduction that Mr. Collins is the most plausible source due to his established tendency toward self-importance and his desire to associate himself with Lady Catherine's authority. This question operates at the Analyze and Evaluate levels. Students must analyze character traits and social relationships. Students must also evaluate which explanation is most consistent with the narrative evidence. Then they must justify their selection based on inference instead of confirmation through the narrative.

This is similar to the thought processes required in film and literary critique. Meaning is often inferred from implication, pattern, and context. The question therefore connects students' narrative comprehension with their interpretive judgment. This demonstrates how MCQs can support the same reasoning processes demanded by formal critique. Other questions in the question bank addressed cinematic form more directly. For example, questions examining camera movement, tone, or performance in the 2005 film adaptation of *Pride and Prejudice* were designed to require students to weigh competing interpretations rather than identify a single correct answer.

- Question:
  - How does the use of long wide-shots in the opening of the 2005 film shape the viewer's understanding of Elizabeth Bennet?

Answer choices were constructed to represent distinct analytical positions justifiable to varying degrees. The task for learners was to determine which interpretation was most strongly supported by cinematic evidence, thereby rehearsing evaluative discrimination instead of recall.

Consistent with best practices in GAI MCQ design, prompts were iteratively refined to improve the cognitive alignment and distractor quality [7, 18]. Initial prompts in the study produced superficial recall questions and answers. Subsequent refinements introduced constraints for the Bloom's level, multiple interpretations, and disciplinary language. One type of prompt used for higher-order items specified:

- The desired Bloom's level.
- The film or literary lens including elements such as cinematography, character interiority, adaptation choice.
- The requirement that more than one answer be defensible.
- The avoidance of signaling a single, unequivocal correct response.

This process reflects a systems-thinking approach to pedagogical design in which the learning objectives and curricular focus as interdependent components that function as the constraints for the prompt engineering process.

2.6 Statistical Framework for Frequency and Item Analysis

A secondary prompt was developed to add a statistical framework for auditing and assessing the quality of the master prompt MCQ output. The design of the prompt was informed by item-writing and item analysis research [7, 14-16, 18]. The MCQs were treated as a dataset to classify across the predefined variables of:

- Presence of item-writing flaws
- Blooms Taxonomy level
- Targeted critique skill
- Single best answer or multiple defensible answer

The system was prompted to produce frequency counts and proportional distributions for each category. The frequency analysis served as a design-level verification mechanism to ensure the intended cognitive distribution and adherence to the item-writing guidelines across the output. Classical psychometric item analysis relies on student response data to calculate perceived difficulty and discrimination processes. However, this prompt acts as a pre-implementation structural audit to assess the MCQs. This approach is consistent with assessment research. ChatGPT was also chosen for this prompt to create consistency for users [18, 20, 33].

3. Results and Discussion

3.1 The Master Prompt

This study produced a primary design artifact in the form of a master prompt for generating a 50-question multiple-choice learning tool for film critique shown in Fig. **1**. This master prompt incorporates scaffolding and constructivist learning by

embedding the critique skill deconstruction, Bloom's Taxonomy, interpretive plurality, and structured output requirements directly into the prompt. The prompt functions as a pedagogical tool that constrains model behavior toward the production of the MCQs, as shown below:

MASTER PROMPT

ROLE:

You are assisting me as an instructional designer. I want to design a 50-question multiple-choice learning tool that teaches students how to think like critics, not simply recall information.

PURPOSE:

- These questions are learning activities, not tests.
- They must train students in the cognitive habits of film and literary criticism.
- Primary text/film: Pride and Prejudice novel, 1995 BBC adaptation, and 2005 film adaptation.
- Target learners: high school or undergraduate students in film, literature, or English.

CRITICAL SKILL FRAMEWORK:

Across the full set, ensure coverage of:

- Perceptual discrimination (noticing framing, sound, tone, rhythm, performance)
- Analytical decomposition (breaking scenes into formal or stylistic elements)
- Interpretive reasoning (connecting form to meaning)
- Comparative judgment (weighing alternative readings or adaptation choices)
- Evaluative justification (arguing why one interpretation is more strongly supported)
- Contextual integration (historical, cultural, ideological awareness)
- Metacognitive awareness (recognizing one's interpretive stance)
- ≥70% must be higher order bloom levels (Analyze/Evaluate/Create)
- 100% of higher-order questions must have multiple defensible answers

STRICT DESIGN RULES:

For Analyze, Evaluate, and Create questions:

- Include at least TWO defensible answer options.
- Avoid a single correct answer.
- Use wording such as most strongly supported, best justified, or most plausible.

For Remember / Understand questions:

- Keep stems short.
- Distractors must be plausible and reflect common misconceptions.
- Purpose is confidence-building and orientation.

Answer choices must:

- Model disciplinary vocabulary where appropriate.
- Represent distinct interpretive positions or reasoning paths.
- Avoid extreme language (always, never, obviously).
- Do NOT invent plot events or film details.
- Maintain a mixed-difficulty rhythm so easier questions are interspersed with harder ones.

CRITICAL SKILL FRAMEWORK: OUTPUT FORMAT (REPEAT FOR EACH QUESTION)

Question #:	Answer analysis:
Question stem	Why A could be defended:
A.	Why B could be defended:
B.	Why C could be defended:
C.	Why D could be defended:
D.	Best-supported option:
Bloom level:	Provide a short evidence-based justification (do not claim certainty).
Critique skill:	Function in sequence:
Lens (e.g., cinematography, tone, performance, narrative, adaptation, symbolism):	Confidence-building Interpretive rehearsal Evaluative judgment

Fig. 1: Master prompt template developed for this study

This master prompt was designed to treat question generation as a design problem. The prompt establishes situational context ("I am an instructor teaching…") and role alignment ("You are assisting me as an instructional designer"). This anchors the

model's responses within an educational design frame. Providing the user identity and task purpose narrows the range of plausible outputs and helps to reduce stylistic drift. The prompt then articulates a clear purpose: the questions are learning activities that teach students how to think like critics. This statement ensures that questions are framed as learning activities instead of defaulting to recall checking.

A strength of the prompt is the inclusion of a deconstruction of film critique skills into perceptual, analytical, interpretive, comparative, evaluative, contextual, and metacognitive components. For a language model, this functions as a conceptual map. It constrains generation toward recognizable cognitive operations. It explains to the system how critique can be understood as a set of learnable micro-skills.

Another strength is in the interpretive plurality rules. Requiring higher-order questions to contain at least two defensible options, framed the answers as comparative judgement tasks. This preserves the interpretive quality of the answers. By requiring Bloom level, critique skill, lens, answer analyses, best-supported option, the prompt transforms each question into a small pedagogical object with metadata. This discourages shallow generation because the model must justify each option and situate it within a cognitive framework.

3.2 Meta-prompting

The master prompt was refined through meta-prompting, the practice of iteratively asking ChatGPT to analyze, diagnose, and strengthen the prompts for its own outputs. Early prompts were simple, such as requests to "create a multiple-choice question for film critique of *Pride and Prejudice*." These produced but shallow recall items. Meta-prompting began by asking: "Why are these questions shallow?" The answer was that the model had not been told what kind of thinking to prioritize. This led to the addition of Bloom-level targeting within the prompt. These steps have been mapped out in Fig. **2**.

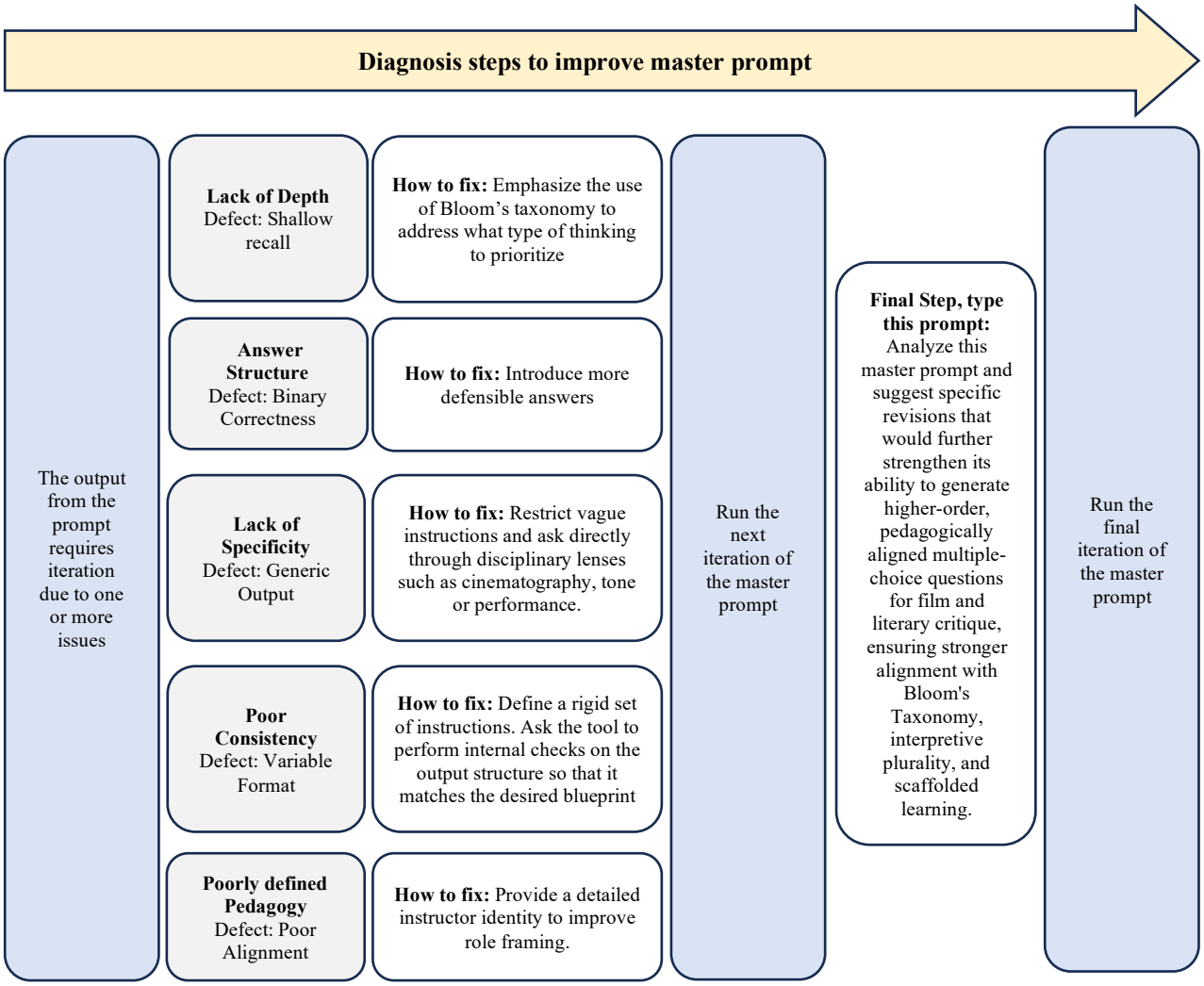

Fig. 2: Steps taken during meta-prompting process to improve prompt accuracy and quality

Further meta-prompting examined why some questions still felt generic. This prompted the addition of the critique skill framework and disciplinary lenses. Instead of asking for "analysis," the prompt now asks for analysis "of cinematography, of performance," etc. Another round of meta-prompting focused on inconsistency across long outputs. This led to the creation of the rigid output template and final reporting section. The meta-prompting question here was: "How can the prompt force consistency without human intervention?" The answer was to require repeated structured fields and a closing audit.

Meta-prompting was also used to address usability by instructors. The prompt was revised to include instructor identity and an instructional designer role as contextual anchoring improves pedagogical alignment. Each refinement step asked "What behavior do I want the system to exhibit, and what instruction would make that behavior unavoidable?" This shift, from content tweaking to behavior shaping, was the defining feature of meta-prompting used in the creation of the master prompt. One final meta-prompt was used once the master prompt was created: "Analyze this master prompt and suggest specific revisions that would further strengthen its ability to generate higher-order, pedagogically aligned multiple-choice questions for film and literary critique, ensuring stronger alignment with Bloom's Taxonomy, interpretive plurality, and scaffolded learning."

AUDIT AND STATISTICAL ANALYSIS

ROLE:

I am conducting a design-level structural audit of these multiple-choice questions for film critique pedagogy. Treat the questions as a dataset. Do not evaluate student performance. Instead, perform a descriptive classification and frequency analysis based on established item-writing guidelines and Bloom's Taxonomy.

FOCAL AREAS:

For each item, classify and report the following:

- Bloom's Taxonomy Level
- Remember
- Understand
- Apply
- Analyze
- Evaluate
- Create
- Targeted Film-Critique Skill
- Perceptual discrimination
- Analytical decomposition
- Interpretive reasoning
- Comparative judgment
- Evaluative justification
- Contextual integration
- Metacognitive awareness
- Item Structure Type
- Single-best-answer
- Multiple-defensible-answer
- Presence of Item-Writing Flaws (based on Haladyna guidelines)

PURPOSE:

Identify whether the item contains any of the following:

- Ambiguous or unfocused stem
- Negative wording
- Implausible distractors
- Unequal option length
- Grammatical cueing
- Logical cueing
- Convergence cues
- "All/None of the above"
- Absolute terms (always, never)

SUMMARY:

After classifying all items:
Generate a structural audit table with the following headings: Item, bloom level, higher order (y/n), critique skill, structure type, flaws present (y/n) and primary flaw (if any).
A. Provide frequency counts for each Bloom level.
B. Provide percentage distribution across Bloom levels.
C. Provide frequency counts for each critique-skill category.
D. Provide counts of single-best vs. multiple-defensible items.
E. Identify any structural imbalances (e.g., overrepresentation of recall-level items).
F. Identify recurring item-writing flaws, if present.
Finally, provide a brief interpretive summary indicating whether the item set reflects:

- Mixed cognitive distribution
- Progressive cognitive scaffolding
- Alignment with higher-order reasoning targets
- Structural weaknesses requiring revision

Fig. 3: Prompt template for auditing the MCQs and performing automated statistical analysis

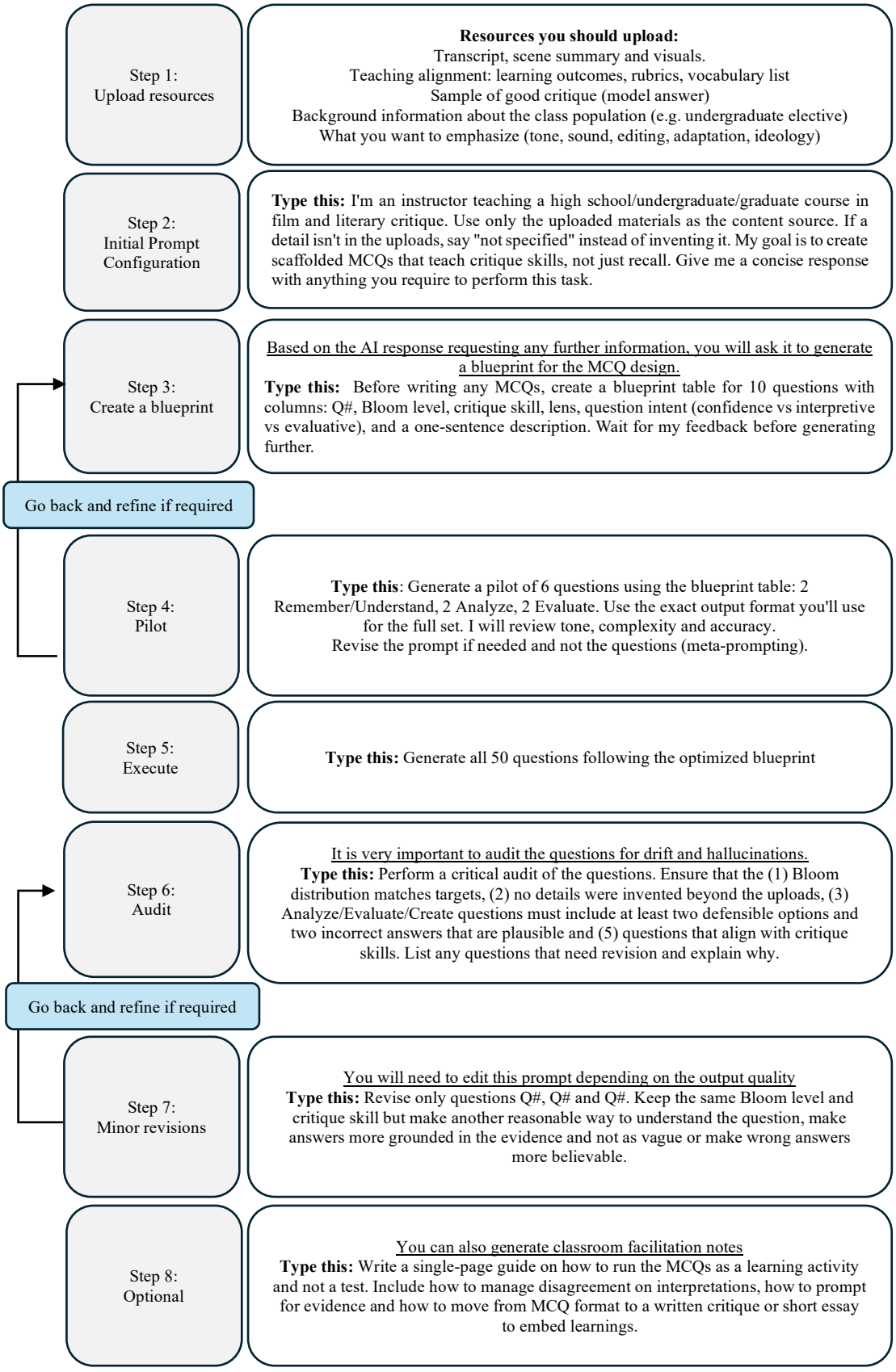

Fig. 4: Overall workflow for MCQ master prompt design, validation and statistical analysis

3.3 Statistical Item and Frequency Analysis Prompt

A secondary prompt was created as a statistical item and frequency analysis audit. This was to create a way to verify the quality of the master prompts outputs. The audit treats the generated MCQs as a structured dataset and classifies each question according to the predefined design variables, including Bloom level, targeted critique skill, item type, and the presence of item-writing flaws. Each question functions as a unit of analysis, allowing frequency counts and proportional distributions to be calculated across categories. The audit prompt produces a structured statistical summary report that includes the frequency and percentage distribution of items across Bloom levels, the distribution of critique skill targets, the counts of single-best versus multiple-defensible-answer items, and the identification of any structural imbalances or recurring item-writing flaws. The report concludes with a brief summary of the results. This process functions as a pre-implementation design-level verification mechanism.

3.4 Guide for Instructors

These frameworks (Fig. 1-4) are intended to be used by instructors with no technical background in GAI or programing presented as a practical, step-by-step guide for instructors on how to use the Master Prompt with ChatGPT to design customized MCQs for film and literary critique. The framework makes the prompt engineering process visible and reproducible. This figure demonstrates what types of materials instructors may upload, what they should communicate to the system at each stage, and how they can iteratively refine outputs to support their pedagogical goals. The framework is not intended as a fixed recipe but as a flexible design process. Instructors are encouraged to adapt, extend, and tweak individual steps based on their individual context. This presents prompt engineering as a form of instructional planning to support instructors in exercising pedagogical judgment. The framework guides instructors in how to interact with GAI through ordinary written instructions, progressively refining those instructions similar to how they would interact with a teaching assistant.

Instructors begin by clearly stating their pedagogical intent in ordinary language. Next, instructors may indicate the level of thinking they want students to practice. Instead of assuming the system understands educational taxonomies, instructors could state the focus, such as basic understanding, analytical comparison, evaluative judgment, or a specific Bloom level. Instructors may then narrow the scope by identifying the specific component of critique they wish to address. This may include cinematography, sound design, performance, narrative structure, character motivation, tone, symbolism, or adaptation choices. Limiting the analytical lens helps prevent generic questions and supports perceptual training by focusing student attention on one formal or thematic element at a time.

Additionally, instructors may clarify how the answer choices should function pedagogically. Rather than serving as correct-versus-incorrect options, answers are designed to represent different interpretive positions. This can include possible but weaker interpretations or common misconceptions. This allows the MCQ format itself to model the required disciplinary thinking. Instructors may also begin with a question they have already written and ask the system to help generate or refine answer choices that align with a particular critique skill. For example, an instructor who has written a question about character motivation might prompt that they want answer options that reflect different interpretations of social behavior and power dynamics. The instructor might ask the system to propose several possible explanations, each grounded in character traits or narrative patterns, and to ensure that more than one option could reasonably be defended using textual or cinematic evidence.

Similarly, if an instructor wishes to focus on cinematography, they might ask the system to generate answer options that correspond to different formal elements, such as the framing, the camera movement. The instructor can then review these options and select or refine those that best model the analytical vocabulary they want students to adopt. When working on tone or mood, an instructor may ask the system to propose interpretations that link lighting, sound, and camera movement to emotional effects. In such questions, students should be required to judge which interpretation is most strongly supported. For questions involving deduction or inference, such as the example involving gossip and character behavior in *Pride and Prejudice*, instructors can explain that the material does not explicitly state the answer and that the goal is to encourage students to infer the most likely explanation based on what is known about character.

The quality and disciplinary specificity of generated questions improves significantly when instructors provide the contextual materials for the GAI system. Uploading items such as short scene descriptions, lecture slides, common interpretations, rubrics, or vocabulary lists helps the prompt generate stronger MCQs. Even brief bullet-point notes about a scene can help anchor the questions in concrete details and reduce the risk of invented or inaccurate outputs.

The examples demonstrate how MCQs can serve as decision-making simulations that help learners practice the analytical and evaluative reasoning involved in film criticism. The study offers a repeatable framework for instructors seeking to utilize GAI to support critique pedagogy.

4. Conclusion

This study addresses a largely overlooked pedagogical challenge in film studies. Students are often expected to perform advanced film criticism without first developing the critical skills necessary for this type of critique. Traditional film pedagogy treats essays, presentations, and discussions as the best way to teach critical thinking. While these formats are essential, they tend to emphasize the final stage of critique rather than the gradual development of perceptual, analytical, and evaluative skills. This research introduces an instructional approach that demonstrates how low-stakes multiple-choice questions can be designed as scaffolded learning activities for learning film critique.

The main contribution of this research is the development of a GAI master prompt that produces film critique MCQs aligned with a deconstructed model of critique skills and Bloom's Taxonomy. This research builds on existing MCQ design that is predominantly focused on medical and nursing training, to provide guidelines for adoption in the humanities. By incorporating cognitive-level distributions, requirements for interpretive plurality, and structured answer rationales, it provides a skill framework especially tailored to film criticism. As a result, this study provides a concrete and reproducible design mechanism for implementing MCQs in film education.

A further contribution of this research is the conceptualization of prompt engineering as a form of pedagogical design. The documented process of meta-prompting illustrates how instructors can progressively shape GAI systems to align with film critique objectives. The master prompt developed in this study enables the generation of higher-order MCQs that serve as interpretive rehearsals for film critique. By requiring multiple defensible options, comparative judgment, and evidence-based justification, these questions simulate the decision-making processes fundamental to film criticism. The inclusion of lower-order MCQs help support student confidence and basic level familiarity with the content.

The study further contributes a statistical item and frequency analysis prompt. This framework can be used for future studies pre- and post-implementation. Post-implementation, the prompt can be customized to include student response data. Distractor selection patterns could provide insight into whether the MCQs effectively scaffold interpretive reasoning. Answers that are consistently answered correctly could be used as recall checks. Unused distractors may indicate insufficient plausibility. This type of analysis could create a feedback loop for refining the prompt constraints. Future research could focus on how students interact with the film critique MCQs over time, assessing the impact of these tools on confidence, interpretive fluency, and writing quality, and the effectiveness of the framework across various films, genres, and educational levels.

This research further addresses a gap in humanities scholarship, which discusses GAI largely in terms of authorship, originality, and ethical considerations instead of instructional design. The study encourages educators to view GAI as a valuable resource for enhancing pedagogical infrastructure. Identifying these opportunities may encourage educators and researchers to further explore and expand this innovative approach. This study demonstrates that, when guided by design constraints and disciplinary intent, GAI can be effectively integrated into film-critique pedagogy to support the development of MCQs to function as micro-sites for critique training within film pedagogy.

Conflict of Interest Statement

The author certifies that there are no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

[1] A. McWhirter, "Film criticism, film scholarship and the video essay," *Screen,* vol. 56, no. 3, pp. 369-377, 2015.

[2] M. Witt, "Taking stock: Two decades of teaching the history, theory, and practice of audiovisual film criticism," 2017.

[3] L. W. Anderson and D. R. Krathwohl, *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc., 2001.

[4] B. BS, "Taxonomy of educational objectives: the classification of educational goals," *Handbook; Cognitive domain,* vol. 1, 1956.

[5] S. Morrison and K. W. Free, "Writing multiple-choice test items that promote and measure critical thinking," *Journal of Nursing Education,* vol. 40, no. 1, pp. 17-24, 2001.

[6] R. E. Tractenberg, M. M. Gushta, S. E. Mulroney, and P. A. Weissinger, "Multiple choice questions can be designed

or revised to challenge learners' critical thinking," *Advances in Health Sciences Education,* vol. 18, no. 5, pp. 945-961, 2013.

[7] S. A. Azer, "Assessment in a problem-based learning course: Twelve tips for constructing multiple choice questions that test students' cognitive skills," *Biochemistry and Molecular Biology Education,* vol. 31, no. 6, pp. 428-434, 2003.

[8] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard university press, 1978.

[9] J. Piaget, "Part I: Cognitive Development in Children--Piaget Development and Learning," *Journal of research in science teaching,* vol. 40, 2003.

[10] D. W. Ecker, "Teaching art criticism as aesthetic inquiry," *Curriculum Theory Network,* vol. 4, no. 2-3, pp. 112-123, 1974.

[11] J. L. Jensen, M. A. McDaniel, S. M. Woodard, and T. A. Kummer, "Teaching to the test… or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding," *Educational Psychology Review,* vol. 26, no. 2, pp. 307-329, 2014.

[12] S. U. Monrad *et al.*, "What faculty write versus what students see? Perspectives on multiple-choice questions using Bloom's taxonomy," *Medical Teacher,* vol. 43, no. 5, pp. 575-582, 2021.

[13] N. L. B. Zaidi *et al.*, "Pushing critical thinking skills with multiple-choice questions: does Bloom's taxonomy work?," *Academic Medicine,* vol. 93, no. 6, pp. 856-859, 2018.

[14] T. Haladyna, "Creating multiple-choice items for testing student learning," *International Journal of Assessment Tools in Education,* vol. 9, no. Special Issue, pp. 6-18, 2022.

[15] T. M. Haladyna, S. M. Downing, and M. C. Rodriguez, "A review of multiple-choice item-writing guidelines for classroom assessment," *Applied measurement in education,* vol. 15, no. 3, pp. 309-333, 2002.

[16] M. Tarrant, A. Knierim, S. K. Hayes, and J. Ware, "The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments," *Nurse Education Today,* vol. 26, no. 8, pp. 662-671, 2006.

[17] Q. Liu, N. Wald, C. Daskon, and T. Harland, "Multiple-choice questions (MCQs) for higher-order cognition: Perspectives of university teachers," *Innovations in Education and Teaching International,* vol. 61, no. 4, pp. 802-814, 2024.

[18] M. E. Magzoub, I. Zafar, F. Munshi, and F. Shersad, "Ten tips to harnessing generative AI for high-quality MCQS in medical education assessment," *Medical Education Online,* vol. 30, no. 1, p. 2532682, 2025.

[19] M. Stadler, A. Horrer, and M. R. Fischer, "Crafting medical MCQs with generative AI: A how-to guide on leveraging ChatGPT," *GMS Journal for Medical Education,* vol. 41, no. 2, p. Doc20, 2024.

[20] K. Hwang, S. Challagundla, M. Alomair, L. K. Chen, and F.-S. Choa, "Towards AI-assisted multiple choice question generation and quality evaluation at scale: Aligning with Bloom's Taxonomy," in *Workshop on Generative AI for Education*, 2023.

[21] H. M. Chan, "Film Criticism in the Digital Age," ed: Taylor & Francis, 2017.

[22] R. K. Anwar, A. A. Mu'yidarrah, and Y. Winoto, "Trends and structures in film criticism research: A bibliometric review of scientific literature," *ProTVF,* vol. 9, no. 2, pp. 170-190, 2025.

[23] A. A. Khan, R. Shabir, and M. Abdullah, "Literature in the Age of AI: How Artificial Intelligence is Reshaping Literary Creation and Criticism," *Journal of Social Sciences and Economics,* vol. 4, no. 1, pp. 77-89, 2025.

[24] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *International Journal of Educational Technology in Higher Education,* vol. 20, no. 1, p. 43, 2023.

[25] A. Mullen, "Reflections on teaching with films in British secondary education," *Journal of Research Initiatives,* vol. 1, no. 3, p. 9, 2015.

[26] N. M. Meer and A. Chapman, "Assessment for confidence: Exploring the impact that low-stakes assessment design has on student retention," *The International Journal of Management Education,* vol. 12, no. 2, pp. 186-192, 2014.

[27] I.Isemonger, "Generative language models in education: Foreign language learning and the teacher as prompt engineer," *TEFL Praxis Journal,* vol. 2, pp. 3-17, 2023.

[28] P. D. Stokes, *Creativity from constraints: The psychology of breakthrough*. Springer Publishing Company, 2005.

[29] G. Sasson Lazovsky, T. Raz, and Y. N. Kenett, "The art of creative inquiry—from question asking to prompt engineering," *The Journal of Creative Behavior,* vol. 59, no. 1, p. e671, 2025.

[30] D. Bordwell, *EBOOK: Film art: An introduction*. McGraw Hill, 2013.

[31] D. Bordwell, "Widescreen aesthetics and mise en scene criticism," *Velvet Light Trap,* vol. 21, pp. 18-25, 1985.

[32] I.Lenchuk and A. Ahmed, "Tapping into Bloom taxonomy's higher-order cognitive processes: the case for multiple choice questions as a valid assessment tool in the ESP classroom," *Arab World English Journal (AWEJ) Special Issue on Covid,* vol. 19, 2021.

[33] D. Kumar, R. Jaipurkar, A. Shekhar, G. Sikri, and V. Srinivas, "Item analysis of multiple choice questions: A quality assurance test for an assessment tool," *Medical Journal Armed Forces India,* vol. 77, pp. S85-S89, 2021.