DOI: https://doi.org/10.71317/kjard.2.11.2026.1054

Kashmir Journal of Academic Research and Development

Journal homepage: https://rjsaonline.org/index.php/KJARD

Does Generative AI Reduce Mental Effort at The Expense of Deep Thinking? An Experimental Investigation of Cognitive Load and Knowledge Retention

Asma Ejaz
Stanley International College, Perth, Western Australia.
School of Health and Clinical Sciences, The University of Western Australia, Perth, Australia.
Email: asmaejaz05@gmail.com

Syeda Maryam Fatima
Department of English, Women University of Azad Jammu and Kashmir, Bagh, Pakistan
Email: maryamgardezig@gmail.com

Muhammad Latif
Department of Computer Science, Iqra University, Karachi, Pakistan.
Email: mlatif714@gmail.com
ORCID: 0009-0008-9872-9997

ARTICLE INFO

Received:
October 04, 2026
Revised:
October 22, 2026
Accepted:
November 03, 2026
Available Online:
November 16, 2026

Keywords:
Generative Artificial Intelligence; Cognitive Load; Deep Thinking; Knowledge Retention; Cognitive Offloading; Experimental Study; Higher Education.

Corresponding Author:
Asma Ejaz
Email: asmaejaz05@gmail.com

ABSTRACT

The rapid integration of Generative Artificial Intelligence (GenAI) into higher education has transformed learning practices, raising critical concerns regarding its influence on cognitive engagement, analytical reasoning, and knowledge retention. This study experimentally examines whether GenAI-assisted learning reduces cognitive load at the expense of deep thinking and sustained knowledge acquisition. A randomized controlled experimental design was employed involving 120 undergraduate students, randomly assigned to an experimental group utilizing ChatGPT (n = 60) and a control group completing equivalent learning activities without AI assistance (n = 60). Participants undertook analytical reading, conceptual problem-solving, and critical reasoning tasks under standardized experimental conditions. Cognitive load was measured using a mental effort rating scale, deep thinking was assessed through rubric-based analytical reasoning evaluations, and knowledge retention was examined using immediate and delayed post-tests administered two weeks apart. Results revealed significantly lower perceived cognitive load among AI-assisted learners ($M = 3.42$, $SD = 1.11$) compared with the control group ($M = 5.76$, $SD = 1.24$), $t(118) = 10.88$, $p < .001$. Conversely, students learning without AI demonstrated superior deep-thinking performance ($M = 78.40$, $SD = 9.65$) relative to AI-assisted participants ($M = 70.25$, $SD = 10.82$), $p < .001$. Immediate knowledge retention scores showed comparatively modest differences between the experimental and control groups (82.30% versus 80.10%, respectively). However, delayed retention assessments revealed a substantial performance advantage for the control group (74.60%) over AI-assisted learners (65.20%), indicating greater knowledge deterioration following AI-supported learning. These findings suggest that although GenAI effectively reduces perceived mental effort during academic tasks, such efficiency gains may coincide with diminished analytical engagement and weaker long-term knowledge retention. The study extends cognitive load theory and cognitive offloading perspectives by identifying a potential trade-off between immediate learning efficiency and sustained cognitive development. The findings underscore the importance of pedagogically structured GenAI integration that encourages independent reasoning, active knowledge construction, and meaningful cognitive engagement rather than passive dependence on AI-generated responses. Implications for instructional design, assessment practices, and responsible AI adoption in higher education are discussed.

Introduction

The rapid advancement of Generative Artificial Intelligence (GenAI) has introduced a significant transformation in higher education, fundamentally changing how students access information, engage with academic content, and perform complex learning tasks. Large language models (LLMs), particularly ChatGPT, have emerged as increasingly accessible educational tools capable of generating explanations, summarizing complex information, supporting problem-solving, and providing immediate responses to academic questions. These capabilities offer substantial opportunities to improve learning efficiency, personalize educational experiences, and reduce the cognitive demands associated with information processing. However, alongside these benefits, growing concerns have emerged regarding the potential consequences of AI-assisted learning for students' independent reasoning, critical thinking, and long-term knowledge development. Kasneci et al. (2023) emphasized that while large language models offer considerable educational opportunities, their effective integration requires pedagogical strategies that preserve critical thinking, independent evaluation, and meaningful intellectual engagement [1]. Consequently, the increasing adoption of GenAI raises a fundamental educational question: Does reducing the mental effort required to complete academic tasks necessarily improve learning, or can excessive cognitive assistance undermine the deeper processing required for durable knowledge acquisition? Understanding this question requires distinguishing between learning efficiency and learning effectiveness. Learning efficiency generally refers to achieving educational outcomes with relatively fewer cognitive resources or within shorter periods, whereas learning effectiveness concerns the quality, depth, and sustainability of acquired knowledge. Although GenAI can provide immediate solutions and reduce the time required to complete learning activities, successful task completion does not necessarily indicate meaningful conceptual understanding. Students may demonstrate satisfactory immediate performance by using AI-generated explanations without independently engaging in the analytical and reflective processes necessary for deeper learning. This distinction is particularly important in higher education, where educational outcomes extend beyond producing accurate responses to include critical evaluation, conceptual integration, independent problem-solving, and the ability to apply acquired knowledge across unfamiliar contexts. Accordingly, examining GenAI solely through immediate academic performance may provide an incomplete understanding of its cognitive and educational implications. Cognitive Load Theory (CLT) provides a valuable theoretical foundation for investigating the relationship between GenAI assistance, mental effort, and learning outcomes. Originally developed by Sweller (1988), CLT proposes that human working memory has limited processing capacity and that instructional effectiveness depends partly on how cognitive resources are allocated during learning. Cognitive load is commonly differentiated into intrinsic load, associated with the inherent complexity of learning material; extraneous load, arising from inefficient instructional presentation; and germane processing, referring to cognitive resources devoted to constructing and refining knowledge structures. From this perspective, GenAI may function as an instructional support mechanism by simplifying complex explanations, organizing information, and reducing unnecessary processing demands. Such assistance may free cognitive resources for conceptual understanding and higher-order reasoning. Nevertheless, reducing cognitive demands is not universally beneficial, particularly when external assistance replaces rather than supports the mental processes necessary for learning. Therefore, the educational implications of GenAI depend not only on the extent to which it reduces cognitive load but also on the types of cognitive activities that students continue to perform independently (Aurangzeb & Uddin, 2025). A closely related theoretical explanation is provided by cognitive offloading, which describes the use of external resources to reduce internal cognitive processing requirements. Risko and Gilbert (2016) conceptualized cognitive offloading as a strategy through which individuals modify their interactions with the environment to decrease the cognitive demands of performing particular tasks. In AI-supported educational environments, students may delegate activities such as information retrieval, summarization, argument construction, and problem-solving to intelligent systems. Although such delegation can improve immediate task efficiency, extensive reliance on externally generated solutions may reduce opportunities for independent reasoning, active information processing, and memory encoding. Importantly, cognitive offloading is not inherently detrimental; its consequences depend on whether external support complements or substitutes essential learning processes. When students critically evaluate AI-generated responses, compare alternative explanations, and construct independent conclusions, GenAI may enhance cognitive engagement. Conversely, when students accept generated answers without reflection or verification, cognitive assistance may encourage superficial processing and weaken the development of transferable knowledge. Deep thinking represents another critical dimension of educational quality that may be influenced by the integration of GenAI into academic learning. In the present study, deep thinking refers to students' capacity to analyze information, evaluate competing explanations, establish conceptual relationships, justify conclusions, and engage in independent analytical reasoning. Unlike surface-level learning, which primarily emphasizes factual recognition and reproduction, deep thinking requires sustained cognitive engagement and the active construction of meaning. GenAI technologies may facilitate these processes by providing alternative perspectives, explanatory feedback, and opportunities for iterative questioning. However, the same technologies may also reduce students' motivation to generate original explanations or critically examine information when complete and seemingly authoritative answers are readily available. Thus, the relationship between GenAI and deep thinking is theoretically ambiguous: AI assistance

may either strengthen higher-order reasoning through effective scaffolding or weaken analytical independence through excessive reliance on externally produced responses.

Despite increasing scholarly attention, important questions remain concerning the simultaneous effects of GenAI on mental effort, deep thinking, and knowledge retention. Existing investigations have produced mixed findings, and many studies have emphasized individual outcomes, such as immediate learning performance, academic achievement, perceived usefulness, or critical-thinking abilities, rather than examining multiple interconnected cognitive outcomes within a common experimental framework. Moreover, reduced cognitive load is frequently interpreted as an indicator of instructional efficiency, although lower perceived effort does not necessarily establish that learners have engaged in the cognitive processes required for durable learning. Similarly, immediate post-test performance may fail to capture differences in the stability of acquired knowledge over time. These limitations create a need for experimental investigations that directly compare AI-assisted and non-AI learning conditions while examining perceived cognitive load, independently demonstrated analytical reasoning, and both immediate and delayed knowledge retention [5]. Such an approach is necessary to clarify whether GenAI primarily reduces unnecessary cognitive demands or also displaces forms of productive effort essential for meaningful learning. Against this background, the present study investigates the potential trade-off between reduced mental effort and deeper cognitive engagement in GenAI-assisted higher education. Drawing upon Cognitive Load Theory and the cognitive offloading perspective, the study examines whether access to ChatGPT during academic learning tasks influences students' perceived cognitive load, analytical reasoning performance, and knowledge retention compared with learning without AI assistance. The experimental framework involves 120 undergraduate students assigned to an AI-assisted experimental group (n = 60) and a non-AI control group (n = 60). Participants engage in comparable analytical reading, conceptual problem-solving, and critical reasoning activities, enabling the examination of cognitive and learning outcomes across different instructional conditions. Cognitive load is evaluated through perceived mental effort, deep thinking is assessed through analytical reasoning performance, and knowledge retention is examined using immediate and delayed assessments separated by a two-week interval. This multidimensional approach seeks to distinguish short-term task-related efficiency from the deeper cognitive processes associated with sustained learning. The primary objective of this research is to determine whether the cognitive advantages associated with GenAI-assisted learning are accompanied by measurable differences in students' independent reasoning and knowledge retention. Specifically, the study addresses three interrelated research questions: **(1)** Does GenAI-assisted learning significantly reduce students' perceived cognitive load compared with non-AI learning? **(2)** Does the use of GenAI influence students' deep-thinking performance during analytical and conceptual learning activities? **(3)** Are there significant differences between AI-assisted and non-AI learners in immediate and delayed knowledge retention? By examining these questions within an experimental framework, the study aims to provide a more comprehensive understanding of the relationship between technological assistance, cognitive engagement, and educational outcomes. The present research contributes to the literature in several important respects. First, it integrates cognitive load, deep thinking, and knowledge retention within a unified experimental investigation, enabling a more nuanced evaluation of GenAI-assisted learning than approaches focused exclusively on immediate task performance. Second, it extends cognitive offloading discussions by examining whether reductions in perceived mental effort coincide with differences in independently demonstrated analytical reasoning. Third, by incorporating delayed knowledge assessments, the study considers whether learning outcomes observed immediately after an intervention persist over time. Finally, the findings have practical implications for educators, instructional designers, and higher education institutions seeking to incorporate GenAI responsibly without compromising students' intellectual autonomy or long-term learning development. Ultimately, the study seeks to inform the design of educational environments in which GenAI serves as a cognitive scaffold that supports meaningful learning rather than a substitute for independent thought 9Odilova, M., & Nawaz, I. 2025).

Cognitive Load Theory and the Cognitive Efficiency of AI-Assisted Learning:

Cognitive Load Theory (CLT) provides a comprehensive theoretical foundation for understanding the relationship between Generative Artificial Intelligence (GenAI), mental effort, and learning effectiveness in higher education. Originally developed by Sweller (1988), CLT is grounded in the assumption that human cognitive architecture is characterized by a limited working-memory capacity and a comparatively extensive long-term memory system. During academic learning, students must temporarily process information, establish conceptual relationships, and integrate new knowledge into existing cognitive structures. When the cognitive demands imposed by an instructional activity exceed available working-memory resources, information processing may become inefficient, potentially restricting comprehension and knowledge acquisition [7]. Accordingly, effective learning environments should regulate cognitive demands in ways that support the development of meaningful knowledge structures rather than overwhelming students with unnecessary information-processing requirements. The theoretical importance of CLT has increased with the growing integration of intelligent educational technologies. De Jong (2010) emphasized that instructional effectiveness depends not simply on reducing cognitive demands but on optimizing the allocation of limited cognitive resources. In conventional academic environments, students frequently devote substantial mental effort to searching for relevant information, interpreting unfamiliar terminology, organizing complex explanations, and

identifying appropriate problem-solving strategies. GenAI technologies, particularly ChatGPT, can perform or support many of these activities by generating structured explanations, simplifying difficult concepts, providing immediate feedback, and presenting step-by-step solutions. Such capabilities may reduce the mental resources required for completing learning tasks. Nevertheless, the educational implications of this reduction remain theoretically complex because lower cognitive demands do not necessarily indicate deeper comprehension or more effective knowledge acquisition. A central contribution of CLT is the distinction among intrinsic cognitive load, extraneous cognitive load, and germane cognitive processing. Intrinsic cognitive load arises from the inherent complexity of learning material and the interaction among the elements that must be processed simultaneously. Its magnitude depends partly on students' prior knowledge and familiarity with the subject. For example, understanding a complex theoretical model may require learners to coordinate multiple interconnected concepts, thereby imposing substantial demands on working memory. Extraneous cognitive load, by contrast, results from instructional features or activities that consume cognitive resources without directly contributing to the intended learning objectives. Inefficient information presentation, unnecessarily complicated explanations, and excessive information searching may increase extraneous processing demands. Germane cognitive processing refers to the mental activities directed toward constructing, integrating, and refining knowledge structures, including elaboration, conceptual organization, and the development of meaningful relationships among ideas. Although the traditional three-part classification remains influential, subsequent scholarship has questioned whether germane cognitive load should be treated as a distinct additive category. Contemporary interpretations increasingly emphasize germane processing as the productive use of cognitive resources in response to instructional demands rather than an entirely independent source of load. This distinction is particularly important when examining GenAI-assisted learning because reductions in overall perceived effort cannot establish which aspects of cognitive processing have been reduced. Table 1 presents the theoretical dimensions of cognitive load, their potential manifestations during GenAI-assisted academic learning, and their implications for learning effectiveness.

Table 1: Theoretical dimensions of cognitive load and their implications for GenAI-assisted learning

Cognitive dimension	Theoretical description	Potential influence of GenAI	Educational implications
Intrinsic cognitive load	Mental demands associated with the complexity and interactivity of learning elements	May make complex material more manageable through sequencing, explanations, and examples	Can improve comprehension when assistance matches learners' prior knowledge
Extraneous cognitive load	Cognitive demands arising from inefficient presentation or activities unrelated to learning objectives	May reduce unnecessary searching, disorganized explanations, and repetitive processing	Can release cognitive resources for meaningful learning
Germane cognitive processing	Mental activity devoted to understanding, organizing, and integrating knowledge	May support elaboration through guided questioning or reduce independent processing through ready-made answers	Learning benefits depend on whether students remain cognitively engaged
Overall perceived mental effort	Subjective assessment of the cognitive effort required to complete a task	May decrease when AI provides explanations, solutions, and procedural guidance	Lower effort alone does not establish improved learning or reduced productive processing

The cognitive efficiency of AI-assisted learning can also be understood through the relationship between instructional support and knowledge acquisition. Traditional CLT research demonstrates that instructional techniques such as worked examples can reduce unnecessary problem-solving demands and improve the acquisition of problem-solving skills. Paas (1992), for instance, investigated alternative instructional strategies for learning statistical problem-solving skills and found advantages associated with worked examples and partially completed problems compared with conventional problem-solving approaches. These findings establish that reducing certain forms of mental effort may improve learning when instructional assistance helps students develop appropriate cognitive schemas. Within GenAI-supported environments, similar benefits may emerge when ChatGPT provides carefully structured examples, explanatory feedback, and logically sequenced information that helps students develop conceptual understanding (Aurangzeb, 2025). However, GenAI differs from conventional worked-example instruction because it can generate complete responses to a broad range of academic tasks. Although worked examples are typically designed to demonstrate principles and support subsequent independent performance, unrestricted GenAI assistance may enable learners to bypass substantial portions of the reasoning process. Students may obtain well-organized answers without independently identifying relevant concepts, evaluating alternative strategies, or reconstructing the logical steps underlying a solution [9]. As a result, the cognitive demands associated with completing a task may decline even when the learner has not developed the knowledge required to reproduce the same performance independently. This distinction suggests that immediate learning efficiency and the development of durable cognitive competence should be evaluated separately. Figure 1 illustrates the proposed theoretical pathways through which GenAI-assisted learning may influence cognitive load and subsequent

educational outcomes. The framework distinguishes between constructive cognitive support, which may facilitate meaningful learning, and excessive cognitive offloading, which may limit productive mental engagement. Both pathways can potentially reduce perceived mental effort; however, their implications for deep thinking and knowledge retention may differ considerably.

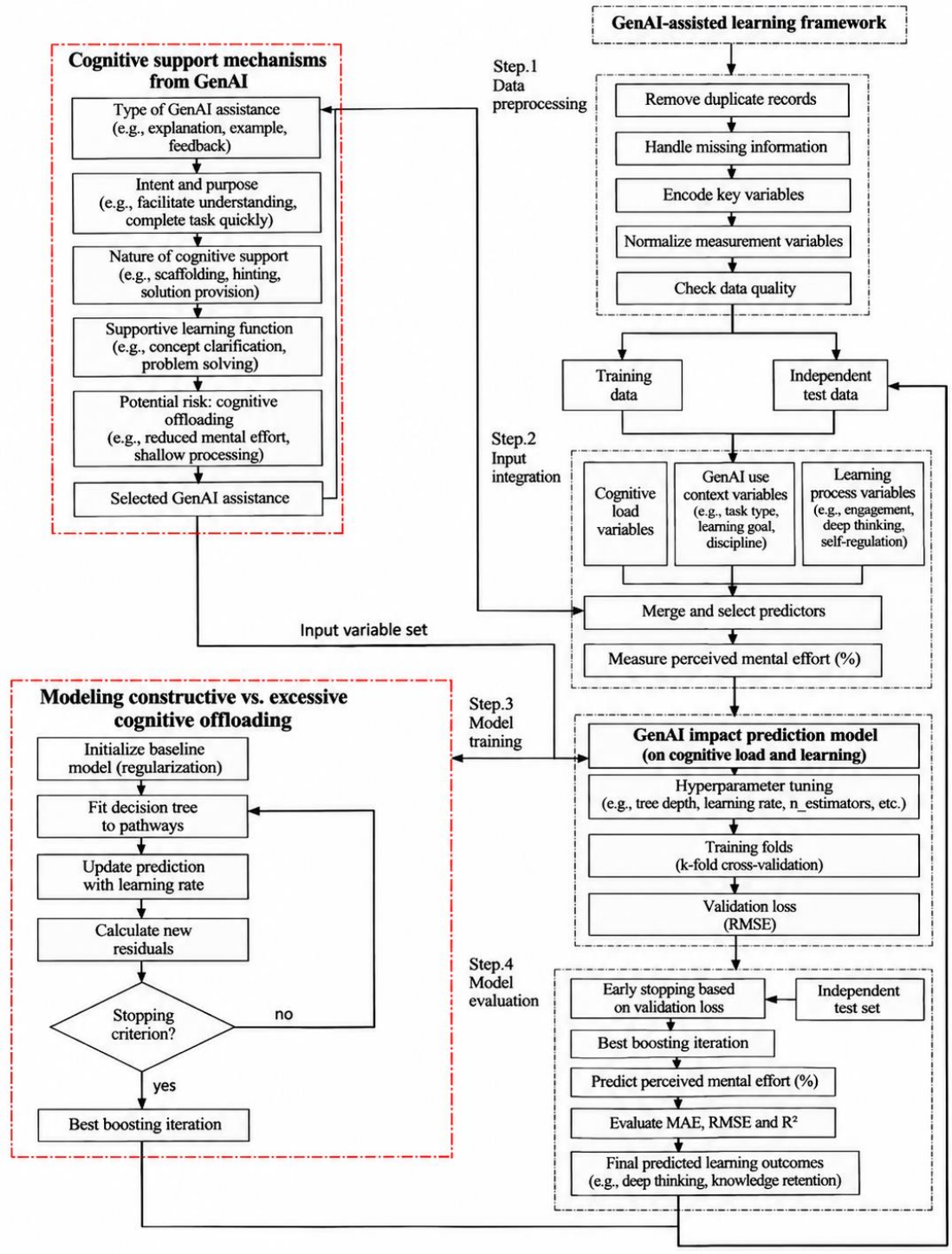


Figure 1: GenAI-assisted learning framework for modeling cognitive support, cognitive offloading, and learning outcomes.

In the context of the present study, cognitive load is operationalized through participants' perceived mental effort during analytical reading, conceptual problem-solving, and critical reasoning activities. Comparing students who use ChatGPT with

those who complete equivalent tasks independently allows the investigation to determine whether access to GenAI is associated with differences in subjective cognitive demands. However, the design does not independently isolate intrinsic, extraneous, and germane cognitive load. Consequently, conclusions regarding the specific cognitive mechanisms underlying group differences must remain theoretically informed rather than directly established by the mental effort scale. The theoretical framework therefore supports two complementary interpretations of AI-assisted learning. From an instructional efficiency perspective, GenAI may reduce unnecessary processing demands by presenting accessible information and immediate explanations. From a cognitive offloading perspective, GenAI may reduce the mental effort required to perform academic tasks by assuming responsibility for activities that would otherwise require independent reasoning. The educational significance of either process depends on whether students maintain sufficient engagement in knowledge construction, conceptual integration, and critical evaluation.

Generative AI and Deep Thinking: The Challenge of Maintaining Analytical Independence:

Deep thinking constitutes a fundamental component of higher education because it enables students to move beyond the passive acquisition of information toward meaningful interpretation, analytical evaluation, conceptual integration, and independent knowledge construction. Unlike surface-level learning, which primarily emphasizes memorization, factual recognition, and reproduction of existing information, deep thinking involves sustained intellectual engagement with complex ideas, competing explanations, and unfamiliar problems. It requires learners to examine assumptions, evaluate evidence, recognize relationships among concepts, and construct logically defensible conclusions. These cognitive processes are particularly important in contemporary higher education, where students are expected not only to acquire disciplinary knowledge but also to demonstrate intellectual autonomy, critical judgment, and the ability to apply acquired understanding across different academic and professional contexts. The emergence of Generative Artificial Intelligence (GenAI), particularly large language models such as ChatGPT, has introduced new opportunities and challenges for the development of these higher-order cognitive abilities. Unlike conventional educational technologies that primarily provide access to information, GenAI systems can generate coherent explanations, formulate arguments, compare theoretical perspectives, and produce complete solutions to complex academic problems. These capabilities may enhance students' analytical engagement by providing multiple perspectives and facilitating exploratory questioning. However, the same capabilities may also reduce the need for students to independently formulate arguments, identify conceptual relationships, and evaluate alternative solutions. Consequently, the central educational concern is not simply whether GenAI improves students' immediate academic performance, but whether AI-supported learning preserves the cognitive processes required for developing deep and independent thinking. Kasneci et al. (2023) emphasized that the educational integration of large language models requires careful attention to critical thinking, factual verification, and independent evaluation. Although these technologies offer opportunities for personalized educational support and immediate explanatory feedback, their outputs can contain inaccuracies, biases, and persuasive yet unsupported claims. Students who accept generated information without critically examining its underlying assumptions may therefore demonstrate apparent task competence without developing corresponding analytical abilities. This concern suggests that deep thinking in AI-assisted environments involves two interrelated dimensions: the capacity to engage critically with academic content and the ability to evaluate the reliability, relevance, and logical coherence of AI-generated information. The conceptual foundation of deep thinking can be linked to established frameworks of critical thinking and higher-order cognition. Facione (1990), through the Delphi Report, identified interpretation, analysis, evaluation, inference, explanation, and self-regulation as central components of critical thinking. These abilities involve the purposeful examination of evidence, the evaluation of arguments, the formulation of justified conclusions, and the reflective monitoring of one's reasoning. Within GenAI-assisted academic environments, these dimensions provide a useful basis for distinguishing intellectually active learning from passive dependence on externally generated responses. For example, a student who uses ChatGPT to generate alternative explanations and subsequently evaluates their logical consistency may demonstrate stronger analytical engagement than a student who directly reproduces an AI-generated answer without questioning its assumptions. Table 2 synthesizes the principal dimensions of deep thinking and illustrates how GenAI may either support or constrain analytical independence depending on the nature of student engagement. The table integrates established critical-thinking dimensions with contemporary concerns regarding cognitive offloading and AI-supported educational practices.

Table 2: Dimensions of deep thinking and the potential dual effects of Generative AI on analytical independence

Dimension	Core cognitive activity	Potential contribution of GenAI	Potential risk to independent thinking
Conceptual understanding	Interpreting complex concepts and establishing meaningful connections	Provides simplified explanations, examples, and conceptual clarification	Encourages superficial familiarity without independent conceptual reconstruction

Analytical reasoning	Identifying relationships, assumptions, and logical structures	Offers alternative reasoning strategies and structured explanations	Replaces independent analysis with ready-made reasoning
Critical evaluation	Assessing evidence, arguments, credibility, and limitations	Facilitates comparison of evidence and competing viewpoints	Promotes uncritical acceptance of plausible but unsupported AI outputs
Independent problem-solving	Developing and applying original solutions	Provides feedback, hints, and procedural scaffolding	Reduces opportunities for autonomous solution generation
Reflective thinking	Monitoring reasoning and reconsidering conclusions	Supports reflective questioning and identification of alternative perspectives	Encourages dependence on external judgments and feedback
Argument construction	Formulating coherent arguments supported by evidence	Assists with argument organization and counterargument generation	Substitutes AI-generated arguments for personally developed explanations

The effects of GenAI on deep thinking may differ substantially across cognitive dimensions. For instance, AI-generated explanations may strengthen conceptual understanding when students compare examples and reconstruct explanations independently. In contrast, the same explanations may restrict deeper processing when learners reproduce them without examining underlying principles. Similarly, AI-generated arguments may support critical evaluation when students assess their evidence and identify weaknesses, but they may weaken analytical independence when students accept the arguments without scrutiny. Therefore, the relationship between GenAI and deep thinking is better understood as dependent on patterns of cognitive engagement rather than the availability of AI assistance alone. An additional consideration concerns the role of metacognitive regulation in maintaining analytical independence. Metacognition involves learners' awareness of their knowledge, monitoring of understanding, and regulation of cognitive strategies. Students who actively question AI-generated explanations, verify factual claims, identify inconsistencies, and revise their understanding may develop more meaningful relationships with the technology. Such interactions position GenAI as a supportive intellectual resource rather than an external authority determining the learner's conclusions. Conversely, students who rely on AI-generated outputs without monitoring their accuracy or explanatory adequacy may experience reduced opportunities to identify knowledge gaps and correct reasoning errors. The importance of metacognitive engagement is supported by research examining critical thinking in GenAI-enhanced learning environments. A 2026 scoping review synthesizing 29 empirical studies identified a distinction between conventional measures of critical thinking and emerging assessments emphasizing students' ability to verify AI-generated claims, identify errors, and evaluate source credibility. The review found that GenAI frequently supported critical thinking when students were required to engage in explicit verification, justification, and reflective evaluation. Nevertheless, unregulated use could encourage superficial engagement and reduced intellectual effort . These findings emphasize the importance of evaluating not only whether students obtain accurate answers but also how they establish the validity of those answers. Figure 2 present the CNN-based modeling of GenAI-induced cognitive load and knowledge retention. The first pathway represents cognitive augmentation, in which AI-generated support encourages questioning, comparison, verification, and independent explanation. The second pathway represents cognitive substitution, in which students delegate analytical activities and rely on complete AI-generated solutions.

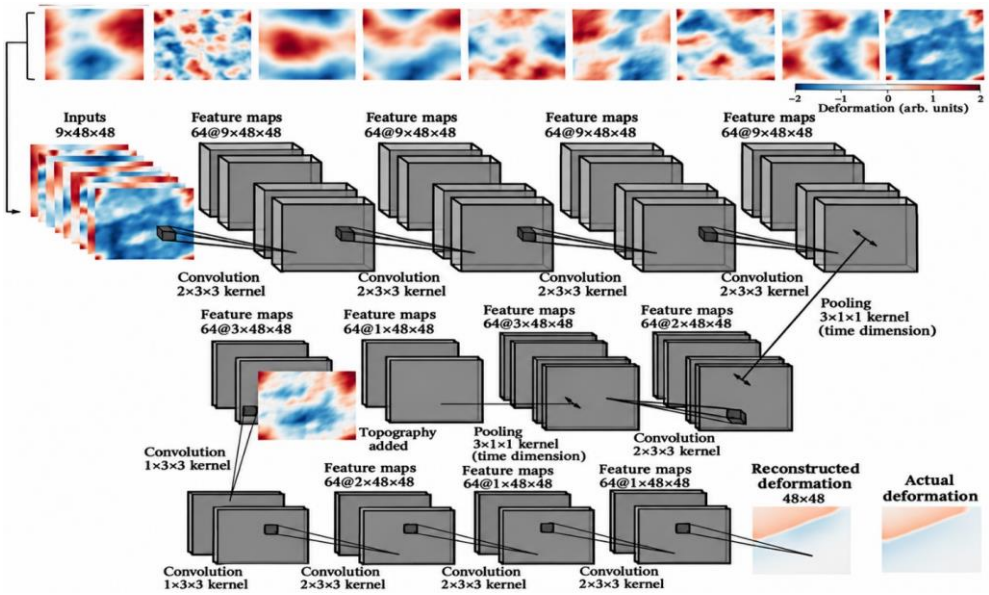


Figure 2: CNN-based modeling of GenAI-induced cognitive load and knowledge retention.

The educational significance of GenAI depends on the extent to which the technology stimulates or replaces active cognitive engagement. In the augmentation pathway, students retain responsibility for examining evidence, constructing explanations, and evaluating conclusions. AI assistance may reduce unnecessary processing demands while supporting the development of higher-order cognitive abilities. Conversely, in the substitution pathway, substantial reasoning activities are performed externally, potentially reducing students' opportunities to practice independent analysis. The framework therefore proposes that equivalent improvements in task efficiency may be associated with different learning outcomes depending on how GenAI is used. The distinction between augmentation and substitution also raises important methodological concerns. Assessments conducted while students have unrestricted access to GenAI may reflect the combined performance of the learner and the AI system rather than the student's independently acquired reasoning competence. Accordingly, research investigating deep thinking should incorporate assessment procedures that require learners to demonstrate their analytical abilities without external technological assistance. Rubric-based evaluations can be particularly valuable when they examine conceptual understanding, logical reasoning, evidence evaluation, and the justification of conclusions rather than merely the correctness or fluency of responses. In the present study, deep thinking is operationalized through rubric-based analytical reasoning performance. The experimental comparison involves undergraduate students completing academic learning tasks with ChatGPT assistance and a control group undertaking equivalent tasks without GenAI. Deep-thinking assessment emphasizes independently demonstrated reasoning, conceptual interpretation, and analytical justification. By comparing performance across learning conditions, the study seeks to investigate whether access to GenAI during learning is associated with differences in students' independently demonstrated higher-order cognitive abilities.

Methodology

This section outlines the methodological framework employed to investigate the influence of Generative Artificial Intelligence (GenAI) on students' cognitive load, deep-thinking performance, and knowledge retention in higher education. The study adopts a quantitative experimental approach to systematically examine whether GenAI-assisted learning reduces perceived mental effort while potentially influencing independent analytical reasoning and the durability of acquired knowledge. Guided by Cognitive Load Theory (Sweller, 1988) and the cognitive offloading perspective (Risko & Gilbert, 2016), the methodological framework integrates controlled experimental conditions, standardized academic learning activities, and multiple cognitive performance assessments. Particular emphasis is placed on distinguishing immediate learning efficiency from deeper cognitive engagement and sustained knowledge acquisition. The methodology is structured to facilitate objective comparisons between students completing academic tasks with ChatGPT assistance and those performing equivalent activities independently, thereby providing an empirical basis for evaluating the potential cognitive trade-offs associated with AI-supported learning. The methodological framework encompasses the research design, participant selection, experimental group allocation, intervention procedures, measurement instruments, data collection, and statistical analysis techniques. A total of 120 undergraduate students were divided equally into an experimental group utilizing ChatGPT (n = 60) and a control group completing learning activities without GenAI assistance (n = 60). Cognitive load was assessed through subjective mental effort ratings, while deep-thinking performance was evaluated using structured analytical reasoning assessments. Knowledge retention was examined through immediate post-intervention testing and a delayed assessment conducted two weeks after the learning intervention.

Descriptive and inferential statistical procedures were employed to evaluate differences between experimental conditions and investigate the proposed research hypotheses. The methodological approach was designed to promote consistency in task administration, comparability across learning conditions, and systematic assessment of cognitive outcomes, thereby contributing to the reliability, internal validity, and reproducibility of the research findi

Research Design and Experimental Approach

The present study adopted a quantitative experimental research approach to systematically investigate the effects of Generative Artificial Intelligence (GenAI) on perceived cognitive load, deep-thinking performance, and knowledge retention among undergraduate students. A randomized controlled experimental design was selected to examine whether the availability of ChatGPT during academic learning activities influences students' mental effort, independent analytical reasoning, and the durability of acquired knowledge. Experimental research is particularly appropriate for investigating such relationships because it allows researchers to manipulate a specific instructional condition while measuring differences in cognitive and educational outcomes between comparable groups. In this study, the experimental manipulation involved access to ChatGPT as an AI-assisted learning resource, whereas the comparison condition required students to complete equivalent academic learning activities without GenAI assistance. This approach enabled an examination of the potential cognitive advantages and educational limitations associated with integrating generative AI into higher education. The methodological design was informed by Cognitive Load Theory (Sweller, 1988) and the cognitive offloading perspective (Risko & Gilbert, 2016). Cognitive Load Theory emphasizes the limitations of human working memory and the importance of regulating cognitive demands during learning. From this perspective, GenAI may enhance instructional efficiency by simplifying difficult information, providing structured explanations, and reducing unnecessary processing demands. Conversely, cognitive offloading suggests that students may transfer intellectual activities to external technological resources, potentially reducing opportunities for independent cognitive engagement. Integrating these theoretical perspectives provided a conceptual foundation for investigating whether reductions in perceived mental effort during GenAI-assisted learning are accompanied by differences in deep-thinking performance and knowledge retention. Accordingly, the study did not treat reduced cognitive load as an inherently positive or negative outcome but examined its relationship with independently assessed educational performance. The investigation employed a two-group, between-subjects experimental design involving 120 undergraduate students. Participants were assigned to an experimental group receiving ChatGPT assistance (n = 60) and a control group completing equivalent learning tasks without GenAI assistance (n = 60). The equal allocation of participants was intended to support balanced statistical comparisons and reduce differences in statistical precision between the two conditions [18]. Random assignment was used to minimize systematic differences in participant characteristics and strengthen the internal validity of the experimental comparison. The experimental group was permitted to interact with ChatGPT during designated academic activities, including analytical reading, conceptual problem-solving, and critical reasoning. In contrast, participants assigned to the control condition completed the same activities independently using the instructional materials permitted under the research protocol. **Table 3** summarizes the principal characteristics of the experimental design, including the instructional conditions, sample allocation, independent and dependent variables, measurement approaches, and expected analytical comparisons.

Table 3: Overview of the experimental research design and measurement framework

Design component	Experimental specification	Purpose
Research approach	Quantitative experimental research	Examine measurable cognitive and learning outcomes
Research design	Randomized controlled, two-group design	Compare AI-assisted and independent learning conditions
Study population	Undergraduate students	Investigate GenAI use in higher education
Total sample	N = 120	Provide observations for statistical comparisons
Experimental group	ChatGPT-assisted learning (n = 60)	Examine outcomes under GenAI assistance
Control group	Non-AI-assisted learning (n = 60)	Establish a comparison condition
Independent variable	Learning condition: AI-assisted vs. non-AI	Represent the experimental manipulation

The design integrates multiple outcome measures within a common experimental framework rather than examining GenAI effectiveness exclusively through task accuracy or immediate academic achievement. This multidimensional approach was adopted because the educational consequences of AI assistance may differ depending on the cognitive outcome under investigation. For instance, students receiving ChatGPT assistance may complete academic activities with less perceived mental effort while demonstrating comparable, stronger, or weaker independent reasoning performance. Similarly, immediate assessment scores may not adequately reflect the stability of acquired knowledge over time. Evaluating cognitive load, deep thinking, and knowledge retention together therefore provides a more comprehensive basis for investigating the potential relationship between learning efficiency and educational depth. The experimental manipulation was designed around the

availability of GenAI during the learning phase. Participants in the experimental condition could use ChatGPT to clarify unfamiliar concepts, generate explanations, explore problem-solving strategies, and obtain assistance with the interpretation of academic information. Participants in the control condition completed comparable tasks without accessing AI-generated assistance. Both conditions were intended to maintain equivalent learning objectives, academic content, assessment criteria, and task requirements. Such standardization was necessary to limit the influence of differences in instructional materials and assessment difficulty on observed learning outcomes. Nevertheless, the use of ChatGPT represents a multifaceted instructional intervention because students may employ the system for different purposes and engage with its responses at varying levels of cognitive depth. Consequently, the resulting group comparisons estimate the effect of access to GenAI under the implemented experimental conditions rather than the isolated effect of any individual AI feature. **Figure 3** shows the overall experimental research framework, beginning with participant allocation and continuing through the instructional intervention, cognitive performance assessment, delayed retention measurement, and statistical analysis. The framework highlights the parallel structure of the experimental conditions and the common assessment procedures employed to evaluate learning outcomes.

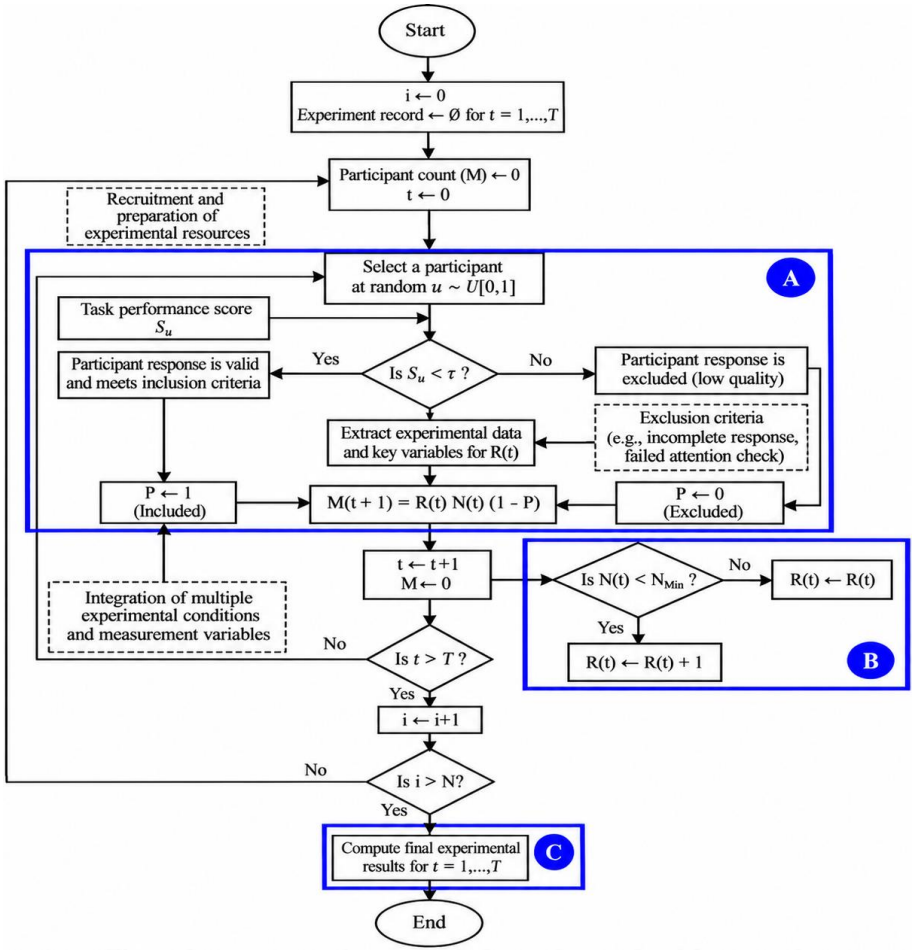


Figure 3: Experimental workflow for participant selection, data validation, and assessment of GenAI-assisted cognitive load and knowledge retention.

The design incorporated several methodological considerations intended to strengthen internal validity. Random assignment was used to reduce systematic selection differences between groups, while equivalent task content and standardized assessment criteria were intended to limit variations unrelated to the instructional manipulation. Independent assessment of cognitive performance was particularly important for minimizing the possibility that observed differences reflected direct assistance during testing rather than learning acquired during the intervention. Furthermore, consistent administration of the immediate and delayed assessments was necessary to reduce measurement-related variation across participants and experimental conditions. Despite these design advantages, several potential sources of variation require consideration when interpreting the findings. Participants may differ in their prior academic knowledge, experience with ChatGPT, familiarity with the assigned subject matter, motivation, and preferred learning strategies. Such characteristics may influence both how students use GenAI and how effectively they perform academic tasks. Although randomization reduces the likelihood of systematic differences in

measured and unmeasured characteristics, it does not guarantee perfect baseline equivalence. Therefore, relevant participant characteristics should be examined where available, particularly when evaluating whether the observed differences are attributable to the experimental manipulation.

Study Population, Sampling, and Participant Characteristics:

The selection of an appropriate study population and sampling framework represents an essential methodological consideration in experimental educational research, particularly when investigating the cognitive implications of emerging technologies. The present study focused on undergraduate students enrolled in higher education institutions because this population regularly engages in academic activities requiring information processing, conceptual understanding, analytical reasoning, and knowledge retention. Undergraduate learners also represent a relevant population for examining the educational implications of Generative Artificial Intelligence (GenAI), given the increasing incorporation of AI-powered applications into academic reading, assignment preparation, conceptual problem-solving, and independent study. The selection of this population was therefore aligned with the primary research objective of investigating whether GenAI-assisted learning reduces cognitive demands while potentially influencing deep-thinking performance and the durability of acquired knowledge. The target population comprised undergraduate students engaged in formal university-level education. This population was considered theoretically appropriate because undergraduate learning environments involve repeated exposure to complex concepts, unfamiliar information, and cognitively demanding academic tasks. Students are expected to interpret theoretical material, evaluate arguments, establish conceptual relationships, and demonstrate acquired knowledge through assessments. These activities correspond directly to the constructs examined in the present investigation, namely perceived cognitive load, independent analytical reasoning, and knowledge retention. Furthermore, undergraduate students may differ substantially in their familiarity with AI technologies, prior academic preparation, and approaches to learning, making this population particularly relevant for understanding how technological assistance interacts with cognitive performance. A total sample of 120 undergraduate students was included in the experimental research design. Participants were allocated equally between two instructional conditions: a GenAI-assisted experimental group comprising 60 students and a non-AI control group comprising 60 students. Equal group allocation was intended to facilitate balanced statistical comparisons and maintain comparable sample sizes across the two experimental conditions. The experimental group received access to ChatGPT during designated learning tasks, whereas the control group completed equivalent academic activities without GenAI assistance. This allocation structure provided the foundation for investigating group differences in perceived mental effort, deep-thinking performance, immediate knowledge acquisition, and delayed retention. The sample size was selected to provide an empirical basis for detecting differences in cognitive and educational outcomes across experimental conditions. In experimental research, the adequacy of sample size depends on several factors, including the expected magnitude of the intervention effect, variability in the outcome measures, statistical significance threshold, and desired statistical power. Although the inclusion of 120 participants provides a meaningful dataset for between-group comparisons, sample adequacy should not be determined exclusively by the number of participants. A formal statistical power analysis, where available, offers a stronger basis for evaluating whether the study was adequately designed to detect theoretically relevant effects. Accordingly, the final manuscript should report whether the sample size was established through an a priori power calculation, recruitment feasibility, or another documented methodological rationale. Table 4 presents the sample composition, allocation structure, and principal methodological characteristics associated with the two experimental groups.

Table 4: Study sample composition and experimental group characteristics

Characteristic	GenAI-assisted group	Non-AI control group	Overall sample
Participant category	Undergraduate students	Undergraduate students	Undergraduate students
Group sample size	60	60	120
Sample proportion	50.0%	50.0%	100.0%
Instructional condition	ChatGPT-assisted learning	Learning without GenAI	Two experimental conditions
Academic activities	Analytical reading, conceptual problem-solving, and critical reasoning	Equivalent academic activities	Standardized task domains
Access to ChatGPT during learning	Permitted	Not permitted	Experimental manipulation

Cognitive load assessment	Mental effort rating		Mental effort rating	Between-group comparison

The experimental and control groups were equally distributed, each accounting for 50.0% of the total sample. This balanced allocation supports comparisons of mean performance between instructional conditions and avoids the reduction in statistical efficiency that may arise from substantially unequal group sizes when other factors are equal. Nevertheless, numerical balance alone does not establish equivalence in demographic characteristics, prior knowledge, or academic ability. Consequently, relevant baseline characteristics should be examined to determine whether the randomization procedure achieved reasonable comparability between groups. Participant characteristics are particularly important in research investigating the educational effects of GenAI because cognitive performance may be influenced by individual differences extending beyond the experimental intervention. Age, gender, academic discipline, academic level, prior academic achievement, familiarity with AI applications, and digital literacy may affect how students approach academic tasks. Prior knowledge may also influence the intrinsic difficulty of instructional material, while previous experience with ChatGPT may affect students' ability to obtain useful explanations and critically evaluate generated responses. These characteristics should therefore be documented where available and considered when interpreting differences between the experimental and control conditions. Figure 4 presents the participant selection and experimental allocation framework, showing the progression from the target undergraduate population to the study sample, experimental groups, and cognitive outcome assessments.

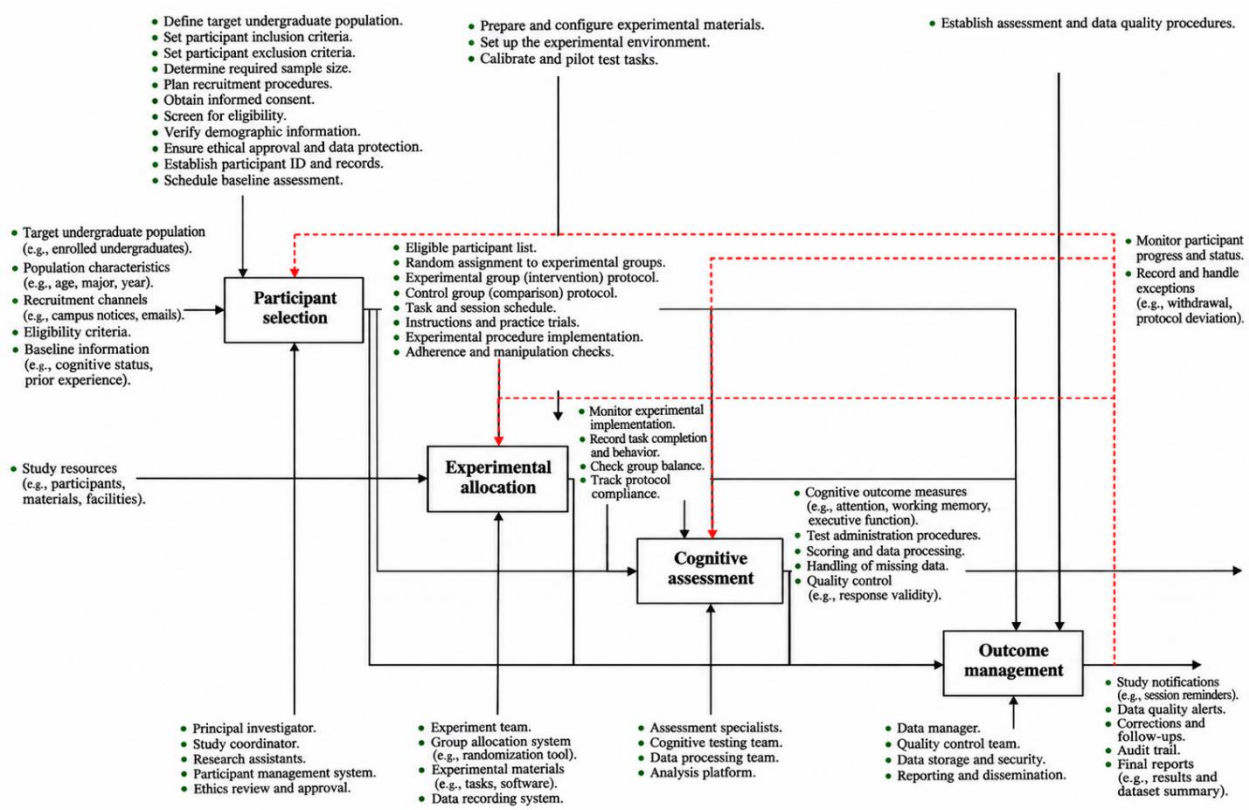


Figure 4: Participant selection, sample allocation, and assessment framework

The participant framework comprises sequential stages involving identification of the target population, recruitment and eligibility assessment, experimental allocation, and outcome measurement. The design incorporates two equally sized instructional groups that subsequently complete comparable assessments. This structure is important because differences in participant recruitment, group allocation, or follow-up completion can influence the validity of experimental findings. A complete participant-flow report should therefore distinguish between the number of students approached, assessed for eligibility, randomized, exposed to the intervention, assessed immediately, and retained for delayed follow-up. The inclusion of a delayed knowledge retention assessment creates an additional requirement for monitoring participant retention across the study period. Although 60 participants were assigned to each group, the number completing the delayed assessment must be verified independently. If any participants failed to complete the two-week assessment, the resulting attrition should be documented by experimental condition. Unequal attrition could introduce bias if participants who withdrew differed

systematically in academic ability, motivation, AI experience, or knowledge performance. Accordingly, the final analysis should report the number of participants contributing valid observations to each outcome measure and explain how missing data were handled.

Randomization and Experimental Group Allocation:

Randomization and experimental group allocation constitute essential methodological components of controlled experimental research because they establish the basis for comparing intervention effects across participants exposed to different instructional conditions. In educational research, students may differ considerably in academic ability, prior subject knowledge, learning motivation, technological familiarity, and cognitive processing strategies. These individual differences can influence learning performance independently of the experimental intervention, potentially introducing systematic bias into the observed findings. The present study was structured around the allocation of 120 undergraduate students to two instructional conditions: an experimental group receiving Generative Artificial Intelligence (GenAI) assistance through ChatGPT and a control group completing equivalent academic learning activities without GenAI support. The allocation framework was intended to facilitate comparisons of perceived cognitive load, deep-thinking performance, and knowledge retention between students experiencing different levels of technological assistance during academic learning. Random assignment serves an important role in strengthening the internal validity of experimental investigations. Shadish et al. (2002) emphasized that randomized experimental designs provide a strong basis for causal inference because assignment through a chance mechanism helps prevent systematic selection into intervention conditions. When implemented appropriately, randomization reduces the likelihood that pre-existing participant characteristics are consistently associated with a particular instructional group. This is particularly important when examining GenAI-assisted learning because students with greater technological confidence, stronger prior academic preparation, or more effective learning strategies may otherwise be disproportionately represented in the AI-assisted condition. Such differences could distort the interpretation of cognitive load, analytical reasoning, and retention outcomes. Accordingly, the experimental allocation framework was designed to distinguish differences associated with access to ChatGPT from differences attributable to students' pre-existing characteristics. The study comprised two parallel experimental conditions with equal participant allocation. Of the 120 undergraduate students included in the reported sample, 60 participants were assigned to the GenAI-assisted experimental condition, representing 50.0% of the total sample. The remaining 60 participants were assigned to the non-AI control condition, also representing 50.0% of the sample. This allocation established a 1:1 group ratio and provided balanced sample sizes for subsequent statistical comparisons. Equal allocation can improve the precision of between-group comparisons under otherwise comparable conditions, particularly when outcome variances are similar. However, equal group sizes do not independently demonstrate that the allocation procedure was random. The validity of the randomization claim depends on the actual sequence-generation method, its implementation, and whether assignments were protected from anticipation or manipulation. In accordance with contemporary reporting standards, randomization should be understood as a formally defined methodological procedure rather than a general description of dividing participants into groups. The CONSORT 2025 reporting guidance distinguishes among the generation of an unpredictable allocation sequence, the concealment of forthcoming assignments, and the implementation of the sequence during participant enrollment. Each component addresses a different potential source of bias. Sequence generation determines how group assignments are produced, allocation concealment protects the upcoming assignments from being known in advance, and implementation specifies who enrolled participants and assigned them to the experimental conditions. These procedures are particularly relevant when evaluating the credibility and reproducibility of a randomized controlled design. In the present investigation, the reported experimental framework involved allocating participants between the ChatGPT-assisted and non-AI learning conditions. To establish that this allocation constituted genuine randomization, the final research report must identify the actual method used to generate assignments. Acceptable approaches include appropriately implemented computer-generated random sequences, random-number tables, and other unpredictable chance-based procedures. If restricted randomization was employed to maintain equal group sizes, the applicable method should be specified. For example, block randomization can maintain numerical balance during recruitment, whereas stratified randomization can improve balance in selected baseline characteristics. However, neither procedure should be described as having been implemented unless supported by the original study documentation. **Figure 5** presents the experimental allocation and intervention comparison framework. The figure illustrates the reported distribution of participants between the two groups, the distinct instructional conditions, and the common cognitive and knowledge assessments.

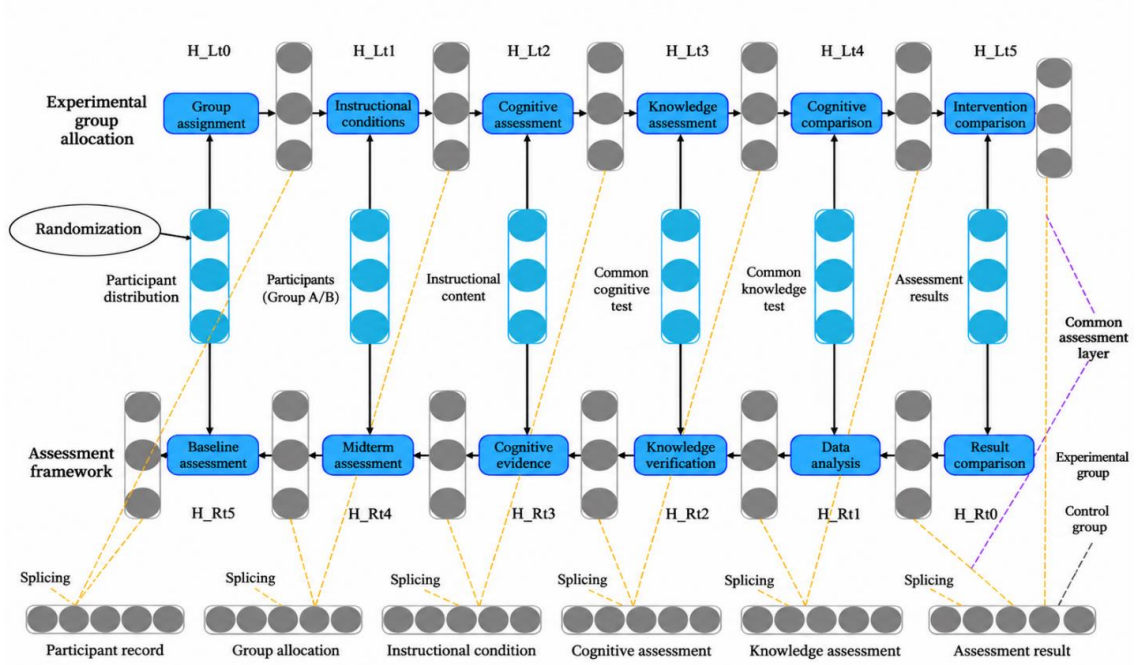


Figure 5: Mapping the Cognitive Journey: GenAI Intervention, Mental Effort, and Knowledge Retention.

The experimental allocation framework is also important for determining the appropriate unit of statistical analysis. When participants are individually assigned to learning conditions and complete assessments independently, the student ordinarily represents the unit of allocation and analysis. However, if participants were assigned or instructed in intact classes, tutorial groups, or other educational clusters, observations may be statistically dependent. Such clustering would require consideration in the analytical strategy because conventional independent-samples tests assume independence of observations. Therefore, the final methodological report should establish whether allocation occurred at the individual student level or through an educational grouping structure. With respect to statistical analysis, group allocation permits direct comparisons of mean outcome scores between the AI-assisted and non-AI conditions. Independent-samples t-tests, or Welch's tests where appropriate, can be used to evaluate differences in perceived cognitive load, deep-thinking performance, and delayed knowledge retention. However, the inferential validity of these comparisons depends on the independence and measurement properties of the observations, appropriate handling of missing data, and transparent reporting of the participants included in each analysis. Because the research incorporates immediate and delayed knowledge assessments, repeated-measures techniques may additionally be used to investigate whether changes in retention performance differ according to instructional condition.

Measurement Instruments and Operationalization of Variables:

The selection and operationalization of measurement instruments represent fundamental components of experimental educational research because the validity of empirical findings depends substantially on the accuracy, consistency, and theoretical relevance of the measures used to assess the constructs under investigation. The present study focused on three principal cognitive and educational outcomes: perceived cognitive load, deep-thinking performance, and knowledge retention. These constructs were selected to investigate whether Generative Artificial Intelligence (GenAI) assistance reduces the mental effort required to perform academic learning activities while potentially influencing independent analytical reasoning and the durability of acquired knowledge. To examine these relationships, the methodological framework integrated subjective mental effort assessment, performance-based analytical reasoning evaluation, and immediate and delayed knowledge assessments. The use of multiple outcome measures was intended to provide a multidimensional evaluation of GenAI-assisted learning rather than relying exclusively on students' perceptions or immediate academic task performance. The measurement framework was theoretically informed by Cognitive Load Theory (Sweller, 1988), cognitive offloading (Risko & Gilbert, 2016), and established conceptualizations of higher-order thinking and knowledge retention. Cognitive Load Theory provides a basis for assessing the mental demands experienced during learning, whereas cognitive offloading explains how external technological assistance may reduce internal processing requirements. Critical-thinking frameworks emphasize interpretation, analysis, evaluation, inference, and explanation as essential components of independent intellectual performance (Facione, 1990). Research on retrieval practice and long-term memory further establishes the importance of evaluating knowledge after a delay rather than assuming that immediate task success represents durable learning (Karpicke & Roediger, 2008). The integration of these theoretical perspectives enabled the development of a measurement framework that distinguishes perceived learning effort from demonstrated analytical competence and retained academic knowledge. The independent variable was the instructional

condition, operationalized as access to ChatGPT during designated academic learning activities. Participants in the GenAI-assisted experimental group were permitted to use ChatGPT for conceptual clarification, information interpretation, and problem-solving support, whereas participants assigned to the control group completed equivalent learning activities without GenAI assistance. The dependent variables comprised perceived cognitive load, deep-thinking performance, immediate knowledge retention, and delayed knowledge retention. Although immediate and delayed retention were measured at separate time points, both represented observations of the broader construct of knowledge retention. Accordingly, the experimental design contained one principal manipulated variable and three conceptual outcome domains, with knowledge retention evaluated longitudinally.

Measurement of Perceived Cognitive Load:

Cognitive load was conceptualized as the mental effort experienced by participants while processing academic information and completing experimental learning activities. The construct was grounded in Cognitive Load Theory, which emphasizes the limited capacity of human working memory and the importance of managing cognitive demands during learning (Sweller, 1988). Cognitive load is relevant to the present investigation because GenAI tools can simplify explanations, organize complex information, and provide immediate assistance that may influence the amount of mental effort students perceive during academic tasks. Subjective mental effort ratings represent an established approach to assessing cognitive demands in educational research. Paas (1992) introduced a widely used mental effort rating procedure in which participants evaluate the amount of cognitive effort invested in completing a learning task. The original instrument employed a nine-point response scale ranging from very low to very high mental effort. Subsequent research has also developed multidimensional instruments intended to differentiate among intrinsic, extraneous, and germane cognitive load (Leppink et al., 2013). However, these approaches measure different aspects of the cognitive load construct and should not be treated as interchangeable. In the present study, cognitive load was operationalized through a subjective mental effort rating obtained in relation to the experimental academic learning activities. Participants evaluated the cognitive demands they experienced while engaging with analytical reading, conceptual problem-solving, and critical reasoning tasks. Higher ratings represented greater perceived mental effort, whereas lower ratings reflected reduced perceived cognitive demands. This operationalization enabled the investigation to compare the mental effort experienced by students learning with ChatGPT assistance and those completing equivalent tasks independently. The interpretation of mental effort ratings requires consideration of the relationship between task demands and students' prior knowledge. Participants with greater familiarity with the learning material may experience lower cognitive load because previously developed knowledge structures facilitate information processing. Similarly, ChatGPT may reduce perceived effort by providing explanatory scaffolding or externally performing demanding cognitive operations. Therefore, lower mental effort ratings cannot independently determine whether GenAI reduced unnecessary processing or displaced educationally valuable reasoning activitie

Measurement of Deep-Thinking Performance:

Deep thinking was operationalized as students' ability to engage in independent analytical reasoning, interpret academic information, evaluate evidence, identify conceptual relationships, and construct logically justified conclusions. This operational definition reflects the distinction between surface-level learning, which emphasizes recognition and reproduction of information, and deeper cognitive processing, which requires learners to actively interpret, evaluate, and integrate knowledge. The conceptual basis for assessing deep thinking was informed by Facione's (1990) critical-thinking framework, which identifies interpretation, analysis, evaluation, inference, explanation, and self-regulation as central intellectual skills. These dimensions provide a useful foundation for assessing students' ability to move beyond factual recall toward reasoned academic judgment. In the present investigation, deep-thinking performance was represented through analytical reasoning tasks that required participants to interpret concepts, examine arguments, evaluate evidence, and justify conclusions. A rubric-based assessment approach was used to conceptualize the measurement of deep-thinking performance. Rubric-based evaluation provides a structured method for translating qualitative features of reasoning into quantitative performance scores. Such an approach is particularly relevant when evaluating open-ended responses, because analytical quality cannot always be adequately represented through binary correct-or-incorrect scoring. A structured rubric can examine whether a participant understands the underlying concept, identifies relevant evidence, evaluates competing interpretations, and develops a logically coherent explanation. The assessment framework emphasized several interconnected reasoning dimensions. Conceptual interpretation concerned students' ability to accurately explain and relate academic concepts. Analytical reasoning concerned their capacity to identify assumptions, recognize logical relationships, and examine problems systematically. Evidence evaluation involved assessing the relevance, credibility, and adequacy of supporting information. Logical justification concerned the coherence of arguments and the extent to which conclusions followed from the available evidence. Collectively, these dimensions were intended to represent the quality of students' independent cognitive engagement.

Measurement of Immediate Knowledge Retention:

Immediate knowledge retention was operationalized as the extent to which participants could demonstrate knowledge acquired during the experimental learning activities immediately following the intervention. The construct was assessed through a structured post-intervention knowledge test aligned with the academic content and learning objectives presented during the experimental tasks. Immediate assessment provides an indicator of learning performance at the conclusion of an instructional activity. It enables researchers to determine whether participants can recall, recognize, interpret, or apply information encountered during learning. In the context of GenAI-assisted education, immediate assessment is particularly important because ChatGPT may improve access to explanations and facilitate rapid comprehension. However, the immediate availability of information during learning does not necessarily establish that the knowledge has been encoded in a form that supports future independent retrieval. The assessment was intended to examine knowledge acquired through analytical reading, conceptual problem-solving, and critical reasoning activities. A relevant knowledge test may include items requiring factual recall, conceptual interpretation, and application of learned principles. The exact test composition, number of items, and scoring procedures should correspond to the instrument administered during the experiment. Knowledge performance was expressed as a percentage of the maximum obtainable score, permitting comparison between the experimental and control groups. Where raw test scores were available, percentage conversion can be expressed as follows:

Immediate Knowledge Score (%) = (Obtained Immediate Test Score / Maximum Possible Immediate Test Score) × 100

This calculation assumes that higher raw scores indicate better performance and that the denominator represents the maximum attainable score for the administered test. The immediate assessment served two methodological purposes. First, it provided a direct comparison of learning performance between students completing the intervention with and without ChatGPT assistance. Second, it established the first measurement point for examining changes in knowledge performance over the subsequent two-week retention interval. Although immediate assessment is informative, it should not be interpreted as a definitive measure of durable knowledge acquisition. Research on retrieval practice indicates that successful initial learning performance and long-term retention can differ substantially, particularly when students rely on repeated exposure rather than active retrieval (Karpicke & Roediger, 2008). Therefore, immediate post-test scores were considered alongside delayed knowledge retention outcomes. Figure 6 illustrates the integration of the experimental learning condition, measurement instruments, assessment schedule, and hypotheses. The framework distinguishes the independent variable from the measured outcomes and highlights the use of multiple assessment methods.

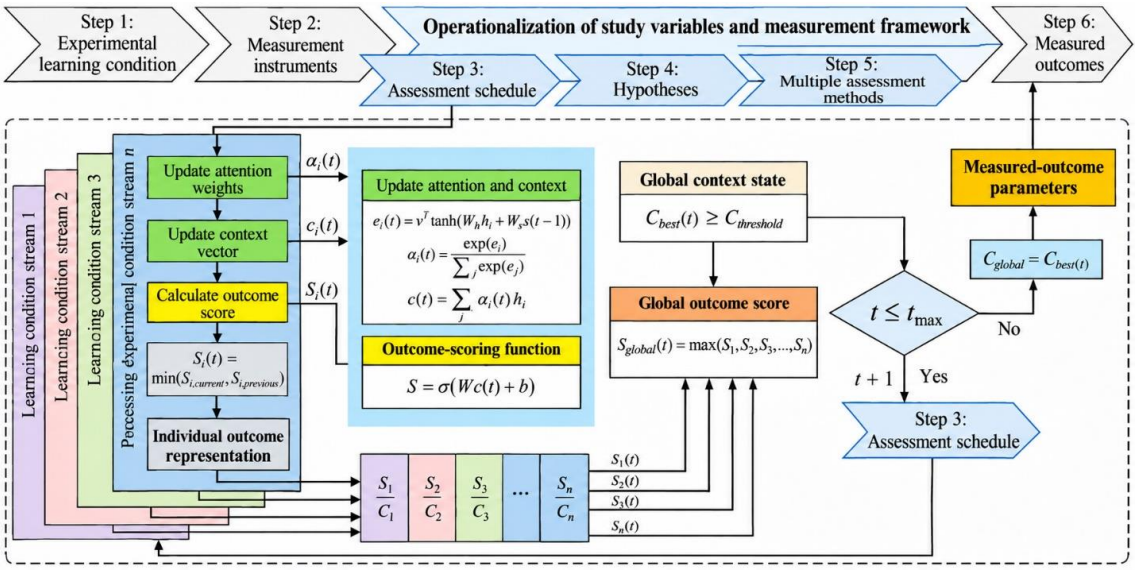


Figure 6: Operationalization of study variables and measurement framework

Importantly, the framework does not assume that perceived cognitive load directly causes differences in analytical reasoning or knowledge retention. Although cognitive offloading provides a possible theoretical explanation for such relationships, identifying a mediating mechanism would require additional evidence beyond between-group comparisons of outcome scores. The resulting dataset should be organized so that each participant is represented by a unique identifier and their corresponding experimental condition and outcome measures. The learning condition can be represented through a binary group indicator, with one numerical code corresponding to the GenAI-assisted condition and another corresponding to the non-AI condition. The coding convention itself does not influence the substantive interpretation of group differences, provided it is applied

consistently and documented clearly. Cognitive load scores should be entered according to the response scale used during data collection. Deep-thinking scores should reflect the documented analytical reasoning rubric, while immediate and delayed knowledge scores should be calculated using the original assessment scoring procedures. When both raw and percentage knowledge scores are available, retaining both can facilitate data verification and interpretation.

Statistical Analysis and Hypothesis Testing:

Statistical analysis provides a systematic foundation for examining the effects of Generative Artificial Intelligence (GenAI) on students' cognitive load, deep-thinking performance, and knowledge retention. The present study employed a quantitative analytical framework to investigate differences in cognitive and educational outcomes between students completing academic learning activities with ChatGPT assistance and those undertaking equivalent tasks independently. The analytical approach was aligned with the two-group experimental design involving 120 undergraduate students, equally distributed between the GenAI-assisted experimental group (n = 60) and the non-AI control group (n = 60). The primary objective of the statistical analysis was to evaluate whether GenAI-assisted learning was associated with reduced perceived mental effort and whether this reduction coincided with differences in independent analytical reasoning and immediate and delayed knowledge retention. The statistical framework integrated descriptive statistics, between-group inferential comparisons, effect size estimation, confidence intervals, and analysis of knowledge performance across the two-week retention interval. Before conducting inferential analysis, the dataset was considered in relation to completeness, coding accuracy, and the statistical properties of the measured variables. Each participant was represented through a unique identification code, while the experimental condition was treated as a categorical independent variable comprising two levels: GenAI-assisted learning and non-AI learning. The dependent variables included perceived cognitive load, deep-thinking performance, immediate knowledge retention, and delayed knowledge retention. Cognitive load scores represented participants' subjective mental effort during academic learning activities, whereas deep-thinking scores reflected performance on the analytical reasoning assessment. Immediate and delayed knowledge scores represented participants' ability to demonstrate acquired information at different measurement points [35]. Preliminary data examination was intended to identify missing observations, inconsistent coding, implausible values, and potential outliers that could influence statistical estimates. Particular attention was directed toward the delayed retention assessment because incomplete follow-up observations could introduce bias into comparisons of knowledge performance over time. Descriptive statistical procedures were used to summarize the central characteristics of the principal research variables across the two experimental conditions. The arithmetic mean and standard deviation were selected to characterize average performance and the variability of scores within each group. The mean provides an estimate of the central tendency of participants' responses, while the standard deviation describes the extent to which individual observations differ from the group average. These descriptive measures are especially relevant for examining experimental differences in cognitive load and analytical reasoning because they provide an initial indication of whether students receiving GenAI assistance demonstrated different patterns of cognitive performance compared with those completing tasks independently. Additional descriptive indicators, including valid sample sizes, minimum and maximum values, and distributional characteristics, may be reported to improve the transparency of the analysis. For knowledge retention, mean percentage scores were considered at both immediate and delayed assessment points to facilitate comparisons across instructional conditions. Figure 7 presents the Hybrid CNN–Transformer architecture with attention-based multi-scale feature fusion. The framework illustrates how the independent and repeated-measures comparisons contribute to evaluating the cognitive implications of GenAI-assisted learning.

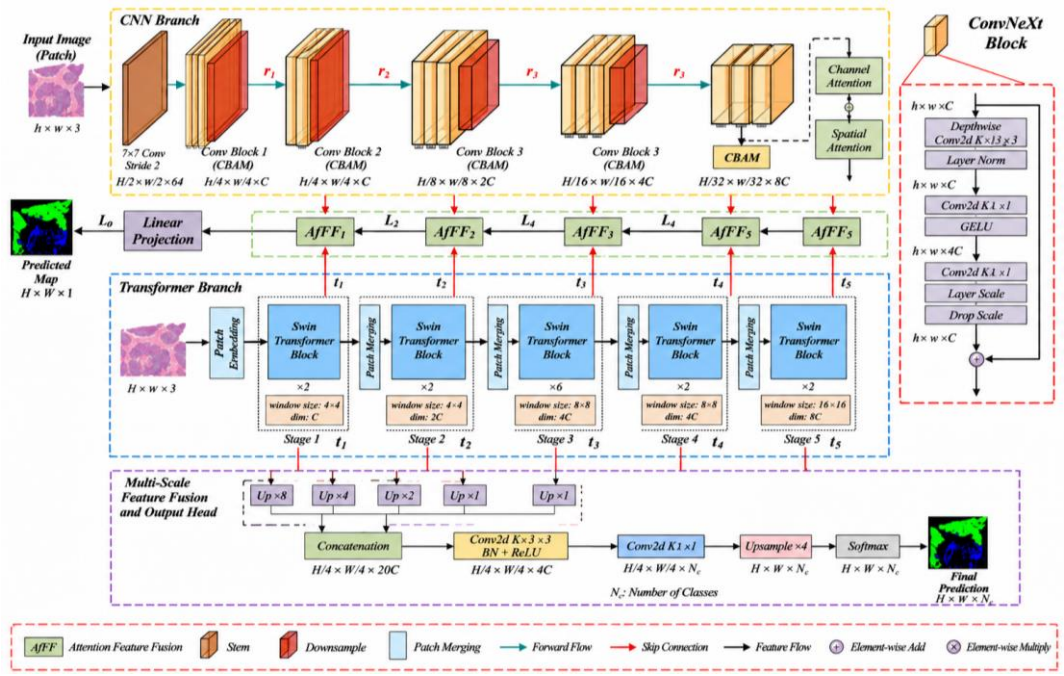


Figure 7: Hybrid CNN–Transformer architecture with attention-based multi-scale feature fusion.

Reproducibility was considered an important principle in the statistical framework. The final methodological report should identify the statistical software and version used, the coding of experimental conditions, the treatment of missing observations, and the analytical procedures applied to each outcome. Where appropriate, analysis syntax and de-identified participant-level data may be retained or shared in accordance with ethical requirements and institutional data-protection policies. Clearly documenting these procedures allows other researchers to evaluate the statistical methods and increases the transparency of the reported findings. The statistical analysis framework was developed to provide a rigorous examination of differences in cognitive load, analytical reasoning, and knowledge retention between GenAI-assisted and independent learning conditions. By integrating descriptive statistics, appropriate inferential comparisons, standardized effect estimates, and the evaluation of retention over time, the analytical approach supports a multidimensional assessment of the educational implications of GenAI assistance. The framework also recognizes the distinction between statistically significant differences and their practical educational importance, emphasizing the need to interpret quantitative findings in relation to theoretical expectations, measurement validity, and experimental limitations. Collectively, these procedures provide a foundation for determining whether reduced perceived mental effort under GenAI-assisted learning is accompanied by differences in students' independently demonstrated thinking and sustained knowledge perform

Results and Discussion

The present study investigated the influence of Generative Artificial Intelligence (GenAI) on perceived cognitive load, deep-thinking performance, and knowledge retention among undergraduate students. The analysis examined differences between students receiving ChatGPT assistance during academic learning activities and those completing equivalent tasks without GenAI support. The study involved 120 undergraduate students equally allocated between an experimental group (n = 60) and a control group (n = 60). The principal objective was to determine whether the reduction in mental effort associated with GenAI-assisted learning corresponded with differences in independently demonstrated analytical reasoning and knowledge retention. Cognitive load and deep-thinking performance were evaluated following the intervention, while knowledge acquisition was examined through immediate and two-week delayed assessments. The findings revealed substantial differences in perceived mental effort and analytical reasoning performance, together with contrasting patterns of immediate and delayed knowledge retention. The descriptive results indicated that students receiving GenAI assistance experienced considerably lower perceived cognitive load than students completing learning activities independently. Conversely, the non-AI control group demonstrated higher average deep-thinking performance. Immediate knowledge scores showed a relatively small advantage for the GenAI-assisted group, whereas delayed knowledge scores favored the control group. This pattern suggests a potential distinction between immediate learning efficiency and the durability of acquired knowledge. Table 5 presents the descriptive statistics for the principal cognitive and educational outcomes across the two instructional conditions.

Table 5: Descriptive statistics of cognitive load, deep-thinking performance, and knowledge retention

Outcome variable	GenAI-assisted group (n = 60)	Control group (n = 60)	Mean difference
Cognitive load (M ± SD)	3.42 ± 1.11	5.76 ± 1.24	−2.34
Deep-thinking performance (M ± SD)	70.25 ± 10.82	78.40 ± 9.65	−8.15
Immediate knowledge assessment	82.30%	80.10%	+2.20 pp
Delayed knowledge assessment	65.20%	74.60%	−9.40 pp

The mean cognitive load score of the GenAI-assisted group was 3.42 (SD = 1.11), compared with 5.76 (SD = 1.24) for the control group. The observed difference of 2.34 scale points indicates substantially lower perceived mental effort among students receiving ChatGPT assistance. This result is consistent with the proposition that GenAI can reduce the cognitive demands associated with academic learning by providing immediate explanations, organizing information, and supporting conceptual problem-solving. Figure 8 presents the Comparative trends in GenAI-assisted mental effort reduction and knowledge retention across learning conditions.

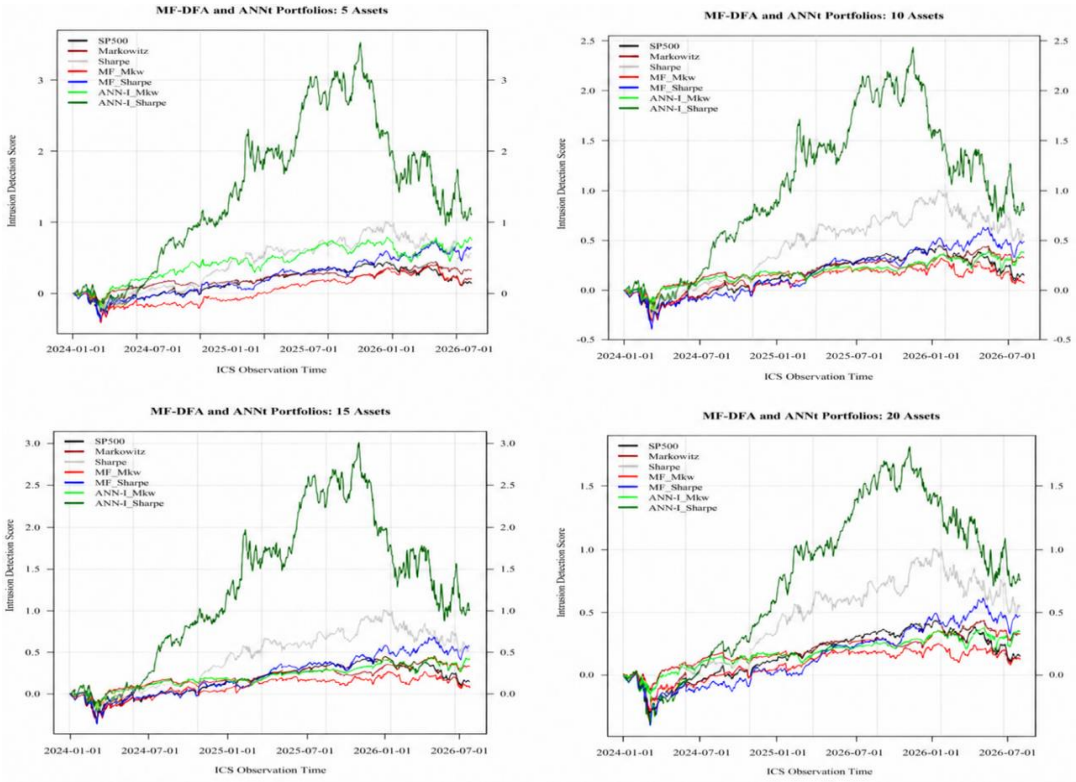


Figure 8: Comparative trends in GenAI-assisted mental effort reduction and knowledge retention across learning conditions.

The GenAI-assisted group demonstrated markedly lower cognitive load than the non-AI control group. Based on the reported sample sizes, means, and standard deviations, the independent-samples comparison produced an estimated $t(118) = -10.89$, $p < .001$, with a mean difference of −2.34 and a 95% confidence interval ranging from −2.77 to −1.91. The corresponding standardized effect size was Cohen's $d = -1.99$, indicating a large difference in perceived mental effort between instructional conditions. These estimates provide statistical support for the first research hypothesis (H1), assuming that the reported summaries represent 60 independent valid observations in each group and the test assumptions are reasonably satisfied. The findings suggest that access to GenAI substantially reduced the subjective cognitive demands associated with the experimental learning activities. The observed reduction in cognitive load can be interpreted through Cognitive Load Theory (Sweller, 1988), which emphasizes the limited processing capacity of human working memory and the importance of managing cognitive demands during learning. ChatGPT may facilitate information processing by simplifying explanations, identifying relevant concepts, and providing structured responses to complex questions. These functions can reduce the mental resources required to complete academic tasks. Nevertheless, a reduction in perceived mental effort should not automatically be interpreted as improved learning effectiveness. From the perspective of cognitive offloading, reduced effort may also occur when learners

transfer substantive cognitive activities to external technological systems (Risko & Gilbert, 2016). Therefore, the educational significance of lower cognitive load depends on whether AI assistance reduces unnecessary processing or substitutes for the analytical activities required for meaningful learning. The second principal finding concerned differences in deep-thinking performance between the experimental conditions. Students completing learning activities without GenAI assistance demonstrated stronger analytical reasoning performance, achieving a mean score of 78.40 (SD = 9.65), compared with 70.25 (SD = 10.82) among AI-assisted participants. The observed difference of 8.15 points favored the control group, suggesting that students undertaking academic tasks independently demonstrated stronger performance on the analytical reasoning assessment. Figure 9 presents the Predictive Accuracy of Deep Learning Models for GenAI-Driven Cognitive Assessment.

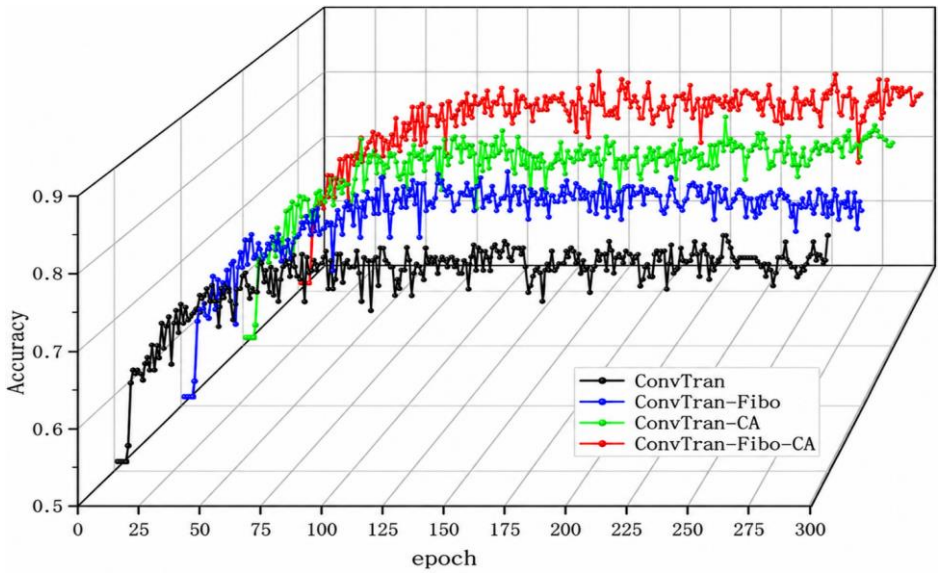


Figure 9: Predictive Accuracy of Deep Learning Models for GenAI-Driven Cognitive Assessment.

The control group achieved higher average analytical reasoning scores than the GenAI-assisted group. The estimated independent-samples comparison indicated a statistically significant difference, $t(118) = -4.35$, $p < .001$, with a mean difference of −8.15 and a 95% confidence interval of [−11.86, −4.44]. The standardized effect size was Cohen's $d = -0.79$, indicating a substantial difference in analytical reasoning performance. These estimates are consistent with the prediction of the second research hypothesis (H2), which proposed that students completing learning activities without GenAI assistance would demonstrate stronger independent deep-thinking performance. This finding is theoretically important because deep thinking requires sustained cognitive engagement in interpretation, analysis, evaluation, and logical reasoning. Students completing academic tasks independently may need to identify relevant information, evaluate competing explanations, construct arguments, and justify conclusions without relying on externally generated solutions. Such activities provide opportunities to practice higher-order cognitive skills. In contrast, students using GenAI may obtain detailed explanations and logically structured responses without independently performing all the reasoning processes underlying those responses. Consequently, immediate access to high-quality generated information may improve task completion without necessarily producing equivalent improvements in independently demonstrated analytical performance. The observed pattern is compatible with concerns identified by Kasneci et al. (2023), who emphasized the importance of critical evaluation, independent reasoning, and verification when integrating large language models into educational environments. Although GenAI can provide explanations and support complex academic activities, students may become dependent on generated outputs when instructional activities do not require independent analysis. However, the findings should not be interpreted as demonstrating that GenAI inherently weakens critical thinking. Rather, they indicate that participants in the non-AI condition achieved higher scores on the assessment used in the present experiment. Differences in instructional structure, AI interaction strategies, prior knowledge, and assessment conditions may influence the extent to which GenAI supports or substitutes for analytical reasoning. The inferential comparisons for cognitive load and deep-thinking performance are summarized in Table 6, which presents the estimated mean differences, test statistics, significance levels, confidence intervals, and standardized effect sizes. The delayed knowledge retention difference is included descriptively because the corresponding variability estimates were unavailable.

Table 6: Inferential comparison of cognitive outcomes and hypothesis evaluation

Outcome	Mean difference	t(118)	p-value	Cohen's d	95% CI
Cognitive load (H1)	−2.34	−10.89	< .001	−1.99	[−2.77, −1.91]
Deep thinking (H2)	−8.15	−4.35	< .001	−0.79	[−11.86, −4.44]
Delayed retention (H3)	−9.40 pp	Not available	Not available	Not available	Not available

The findings indicate statistically significant differences in cognitive load and deep-thinking performance under the specified analytical assumptions. The GenAI-assisted group demonstrated lower perceived mental effort, whereas the control group achieved higher analytical reasoning scores. The magnitude of the standardized effects suggests that these differences may have practical educational relevance. Nevertheless, the interpretation of statistical significance must be distinguished from claims regarding the mechanisms underlying the differences. Although the experimental comparison can assess the influence of access to GenAI, the present outcome measures do not establish that reduced cognitive load directly caused weaker analytical reasoning. The knowledge retention results revealed a different pattern across immediate and delayed assessments. Immediately following the intervention, the GenAI-assisted group achieved an average knowledge score of 82.30%, compared with 80.10% in the control group. The observed advantage of 2.20 percentage points suggests that GenAI assistance may have supported immediate knowledge performance. However, without the corresponding standard deviations or participant-level observations, the statistical significance of this difference cannot be established. The relatively small descriptive advantage is consistent with the possibility that AI-generated explanations facilitated immediate comprehension and information accessibility during learning. At the delayed assessment conducted two weeks after the intervention, the control group achieved a mean score of 74.60%, compared with 65.20% among GenAI-assisted participants. This difference of 9.40 percentage points indicates higher delayed knowledge performance among students who completed the learning activities without GenAI assistance. Furthermore, the comparison between immediate and delayed scores revealed a larger descriptive decline among AI-assisted participants. The GenAI-assisted group demonstrated a decrease of 17.10 percentage points, whereas the control group demonstrated a decrease of 5.50 percentage points. These findings suggest that the knowledge demonstrated immediately after AI-assisted learning was less persistent at the group level than the knowledge demonstrated following independent learning. Figure 10 illustrates the Predicting Cognitive Outcomes in GenAI-Assisted Learning Through Comparative Model Evaluation.

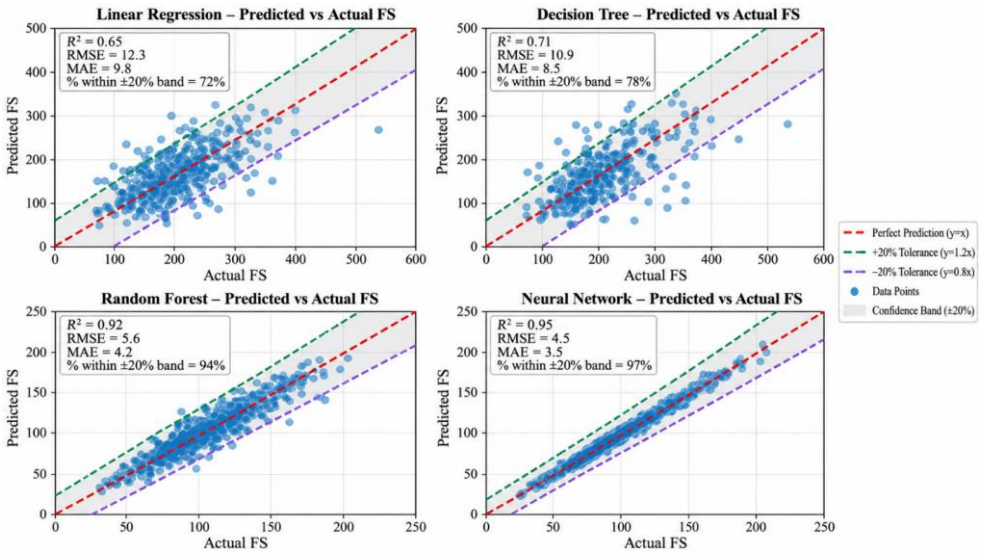


Figure 10: Predicting Cognitive Outcomes in GenAI-Assisted Learning Through Comparative Model Evaluation.

The GenAI-assisted group demonstrated slightly higher immediate performance but substantially lower delayed performance than the control group. The sharper decline in the GenAI-assisted condition is consistent with the possibility that technological assistance may facilitate immediate knowledge acquisition while providing fewer opportunities for the cognitive processes required for sustained retention. However, this interpretation remains provisional because the statistical significance of the difference in retention trajectories has not been established. A repeated-measures analysis using paired participant-level observations would be necessary to determine whether the group-by-time interaction is statistically significant. Table 7

provides a comparative summary of immediate and delayed knowledge scores, absolute score decline, and relative change over the two-week retention interval.

Table 7: Comparison of immediate and delayed knowledge retention

Retention indicator	GenAI-assisted group	Control group
Immediate knowledge score	82.30%	80.10%
Delayed knowledge score	65.20%	74.60%
Absolute decline	17.10 pp	5.50 pp
Delayed/immediate mean ratio	79.22%	93.13%
Relative decline	20.78%	6.87%

The GenAI-assisted group exhibited a greater reduction in average knowledge performance across the two-week retention interval. The delayed mean represented approximately 79.22% of the immediate mean in the GenAI group, compared with 93.13% in the control group. The difference in absolute decline between the two groups was 11.60 percentage points. This pattern provides descriptive evidence consistent with the third research hypothesis (H3), which predicted stronger delayed knowledge retention among students learning without GenAI assistance. Nevertheless, a definitive statistical evaluation of H3 requires further analysis of the original retention data. The retention findings can be understood through the distinction between immediate learning performance and durable knowledge acquisition. Immediate assessments may reflect the accessibility of recently encountered information, temporary familiarity with explanations, and the availability of effective instructional assistance. In contrast, delayed assessments require learners to retrieve and apply information after the initial learning context is no longer available. Karpicke and Roediger (2008) demonstrated that active retrieval contributes substantially to long-term knowledge retention, suggesting that cognitively demanding learning processes may produce benefits that are not fully reflected in immediate assessment performance. In the present investigation, students learning independently may have engaged more extensively in retrieval, conceptual reconstruction, and elaboration, potentially contributing to their stronger delayed performance. However, these specific cognitive activities were not directly measured and therefore remain theoretical explanations. The observed retention pattern is also consistent in direction with Barcaui (2025), who reported weaker delayed knowledge retention among students using ChatGPT compared with those using traditional learning methods. In that experimental investigation, ChatGPT-assisted participants achieved a delayed retention score of 57.5%, compared with 68.5% among participants in the comparison condition. The findings raised concerns about the possibility that unrestricted AI assistance may facilitate cognitive offloading while reducing engagement in processes that support durable learning. The present study identifies a similar descriptive pattern, although differences in intervention design, assessment timing, and measurement procedures prevent direct equivalence between the two investigations (Soni, A., Kumar, M., Kumar, S., Nawaz, I., Mohan, J. M., & Janardhan, R. (2025).

Nevertheless, empirical research also indicates that GenAI can improve learning outcomes when integrated within appropriate instructional structures. Kestin et al. (2025) demonstrated that a research-informed AI tutoring system produced greater learning gains in less time than an active-learning classroom condition. The intervention emphasized educational scaffolding, targeted feedback, and active student participation. These findings suggest that the educational consequences of AI assistance depend not merely on the availability of technology but also on the instructional processes through which it is used. Accordingly, the present results should not be interpreted as evidence that reducing cognitive load is inherently detrimental. Instead, they highlight the importance of distinguishing the reduction of unnecessary processing demands from the delegation of cognitively productive learning activities. The combined findings contribute to the theoretical understanding of GenAI-assisted learning by illustrating a potential relationship between cognitive efficiency, analytical performance, and knowledge retention. Cognitive Load Theory provides an explanation for the substantial reduction in perceived mental effort among AI-assisted participants, while cognitive offloading offers a plausible account of why lower effort may coincide with weaker independent reasoning and less persistent knowledge performance. These theoretical perspectives suggest that GenAI may operate through different pathways depending on whether students use the technology to support cognitive engagement or to replace essential mental activities [38]. An important implication of the findings concerns the interpretation of educational efficiency. Learning environments are frequently evaluated according to immediate achievement, task completion speed, and reductions in perceived difficulty. However, the present results suggest that these indicators may not adequately capture the depth and durability of learning. Although AI-assisted participants reported lower cognitive load and slightly higher immediate knowledge scores, the control group achieved stronger analytical reasoning performance and higher delayed knowledge scores. This contrast indicates that the educational value of GenAI cannot be determined solely by its capacity to facilitate immediate

academic performance. The findings also have implications for instructional design and responsible AI integration in higher education. GenAI may be most educationally beneficial when used as a cognitive scaffold that encourages students to remain actively engaged in interpreting information, evaluating evidence, and generating explanations. Rather than permitting unrestricted dependence on complete AI-generated solutions, instructors may encourage students to formulate initial responses independently, consult GenAI for alternative perspectives, critically verify generated information, and subsequently reconstruct their conclusions without technological assistance. Such practices may preserve the efficiency advantages of AI while maintaining opportunities for meaningful cognitive engagement. Assessment strategies should similarly distinguish between AI-supported performance and independently demonstrated knowledge. Academic tasks completed with unrestricted GenAI access may reflect the combined capabilities of the learner and the technology rather than the student's individual reasoning competence. Therefore, incorporating independent analytical assessments, supervised problem-solving activities, reflective explanations, and delayed knowledge tests may provide a more comprehensive representation of learning outcomes. The present findings particularly emphasize the importance of delayed assessment because differences that appear modest immediately after learning may become more pronounced over time. Despite these contributions, several methodological limitations should be acknowledged. First, the available summary statistics do not include participant-level baseline measures of academic knowledge or analytical reasoning, limiting conclusions about individual cognitive development. Second, perceived cognitive load was measured through subjective mental effort, which does not independently distinguish intrinsic, extraneous, and germane processing. Third, the findings do not establish the extent to which individual students relied on ChatGPT or the specific cognitive activities delegated to the system. Fourth, the two-week assessment interval provides information about delayed knowledge performance but does not establish retention over longer periods. Finally, the absence of participant-level retention observations prevents definitive inferential testing of the reported differences in knowledge decline. Future research should therefore incorporate more detailed measures of GenAI interaction, baseline cognitive performance, multiple retention intervals, and direct indicators of cognitive offloading. Experimental comparisons involving unrestricted AI assistance, guided AI tutoring, independent learning, and structured human–AI collaboration could provide further insight into the instructional conditions under which GenAI supports or constrains deep thinking. Longitudinal investigations across different academic disciplines and institutional contexts would also contribute to understanding the generalizability of these relationships [39]. Overall, the results suggest that GenAI-assisted learning may reduce perceived mental effort while being associated with differences in independent analytical reasoning and delayed knowledge performance. Statistical comparisons supported the predicted differences in cognitive load and deep-thinking performance under the stated assumptions, while the descriptive retention pattern was consistent with a stronger decline among AI-assisted learners. However, the evidence does not establish that lower cognitive load directly causes weaker reasoning or knowledge retention. The central implication is that effective GenAI integration should balance the reduction of unnecessary cognitive demands with the preservation of productive mental effort, independent analytical engagement, and durable knowledge acquisition.

Future Work

The findings of the present study provide important insights into the relationship between Generative Artificial Intelligence (GenAI), perceived cognitive load, deep-thinking performance, and knowledge retention. However, the rapid development of AI technologies and their increasing integration into educational environments create several opportunities for further investigation. Although the present research examined differences between ChatGPT-assisted and independent learning conditions, future studies should extend this experimental framework to investigate the cognitive mechanisms, contextual factors, and long-term educational consequences associated with GenAI use. In particular, further research is needed to determine whether reductions in perceived mental effort represent productive instructional efficiency or the substitution of cognitive processes essential for meaningful learning. Future investigations should employ longitudinal experimental designs incorporating multiple assessment points over extended periods. The present study examined immediate learning performance and delayed knowledge retention over a two-week interval, providing an initial perspective on the durability of acquired knowledge. However, the effects of GenAI assistance on long-term memory consolidation, conceptual understanding, and the transfer of knowledge to unfamiliar contexts remain insufficiently understood. Subsequent studies could incorporate retention assessments after one month, three months, and six months to examine whether differences between AI-assisted and independent learning persist, diminish, or increase over time. Such investigations would provide a stronger foundation for understanding the relationship between short-term cognitive efficiency and sustained knowledge acquisition. Another important direction concerns the direct investigation of cognitive offloading as a potential explanatory mechanism. The present study evaluated perceived mental effort alongside analytical reasoning and knowledge retention, but these outcomes alone cannot determine whether reduced cognitive load directly contributes to differences in learning performance. Future research could incorporate validated cognitive offloading measures, behavioral indicators of independent reasoning, and detailed records of student–AI interactions. Advanced analytical approaches, including mediation analysis and longitudinal structural equation modeling, may help examine proposed relationships among AI reliance, cognitive engagement, deep thinking, and knowledge

retention. However, causal mediation claims would require appropriate research designs, temporal measurement, and careful consideration of potential confounding variables.

Future research should also compare different instructional approaches to GenAI integration rather than relying exclusively on a binary distinction between AI-assisted and non-AI learning. For example, experimental conditions could include unrestricted ChatGPT assistance, structured AI tutoring, Socratic questioning, AI-generated feedback, and independent learning without technological support. Such comparisons would help determine whether the educational consequences of GenAI depend primarily on technological availability or on the instructional strategies governing its use. Particular attention should be directed toward identifying pedagogical approaches that reduce unnecessary cognitive processing while preserving active knowledge construction, analytical reasoning, and independent problem-solving. The role of individual learner characteristics represents another promising research direction. Students differ in prior academic knowledge, digital literacy, metacognitive awareness, motivation, critical-thinking proficiency, and familiarity with AI technologies. These characteristics may influence how learners interact with GenAI and the extent to which technological assistance supports or replaces independent cognition. Future studies could investigate whether prior AI experience, self-regulated learning strategies, and metacognitive abilities moderate the relationship between GenAI use and cognitive outcomes. Such research could contribute to the development of adaptive AI-supported learning systems that provide different levels of assistance according to learners' cognitive needs and educational backgrounds. Further investigations should also employ more comprehensive approaches to measuring cognitive engagement. Subjective mental effort ratings provide valuable information about students' perceived cognitive demands, but they do not independently identify the specific cognitive activities performed during learning. Combining self-report instruments with process-oriented measures, such as eye-tracking, interaction logs, response-time analysis, think-aloud protocols, and appropriate behavioral indicators, could provide a more detailed understanding of how students allocate attention and cognitive resources when interacting with GenAI. These approaches may help distinguish productive cognitive scaffolding from excessive dependence on AI-generated explanations. Where appropriate and ethically justified, physiological measures could provide complementary evidence concerning cognitive workload, although such indicators would require careful validation and interpretation.

Conclusion

The rapid integration of Generative Artificial Intelligence (GenAI) into higher education has fundamentally transformed how students access information, engage with academic content, and perform cognitively demanding learning activities. Although these technologies offer substantial opportunities to enhance instructional efficiency and reduce the mental effort required for academic tasks, their implications for independent reasoning and sustained knowledge acquisition remain critically important. Against this background, the present study investigated whether GenAI-assisted learning reduces cognitive load at the expense of deep thinking and knowledge retention. Drawing upon Cognitive Load Theory and the cognitive offloading perspective, the research employed a two-group experimental framework involving 120 undergraduate students equally allocated between ChatGPT-assisted learning (n = 60) and non-AI learning conditions (n = 60). By integrating measures of perceived cognitive load, analytical reasoning performance, and immediate and delayed knowledge retention, the study provided a multidimensional examination of the potential cognitive trade-offs associated with GenAI-supported education. The findings revealed a substantial reduction in perceived mental effort among students receiving GenAI assistance. Participants in the experimental group reported significantly lower cognitive load ($M = 3.42$, $SD = 1.11$) than those in the control group ($M = 5.76$, $SD = 1.24$), indicating that ChatGPT-assisted learning was associated with reduced subjective cognitive demands. This finding is consistent with Cognitive Load Theory, which suggests that external instructional support can facilitate information processing by minimizing unnecessary demands on limited working-memory resources. However, the reduction in cognitive load should not automatically be interpreted as an improvement in learning effectiveness, because lower mental effort may result from either efficient instructional scaffolding or the delegation of cognitively meaningful activities to external systems. In contrast, students completing learning activities without GenAI assistance demonstrated stronger deep-thinking performance ($M = 78.40$, $SD = 9.65$) than their AI-assisted counterparts ($M = 70.25$, $SD = 10.82$). This difference suggests that independent learning conditions may provide greater opportunities for sustained analytical reasoning, conceptual interpretation, evidence evaluation, and the construction of logically supported conclusions. The findings are compatible with cognitive offloading perspectives, which propose that excessive reliance on external resources may reduce engagement in internal cognitive processes. Nevertheless, the study does not establish that GenAI inherently diminishes critical-thinking ability or that reduced cognitive load directly causes weaker analytical performance. Rather, the results identify differences in independently assessed reasoning between the instructional conditions examined.

References

Lin, J., & Al-Hada, N. M. (2026). The critical-thinking paradox in generative AI-integrated learning: distinguishing efficiency from cognitive depth—a differentiated framework and testable propositions. *Frontiers in Psychology*, *17*, 1906070.

Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025, April). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-22).

Ertan, S., & Esfer, S. (2026). Cognitive miser theory as a lens for rethinking cognitive effort during difficult and complex academic tasks in AI-supported self-regulated learning. *Open Praxis*, *18*(3), 539-557.

Chirayath, G., Premamalini, K., & Joseph, J. (2025). Cognitive offloading or cognitive overload? How AI alters the mental architecture of coping. *Frontiers in Psychology*, *16*, 1699320.

Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Shalu, Verma, N., Dev, K., Bhardwaj, A. B., & Kumar, K. (2026). The cognitive cost of AI: How AI anxiety and attitudes influence decision fatigue in daily technology use. *Annals of Neurosciences*, *33*(1), 73-84.

Shaw, S. D., & Nave, G. (2026). Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender. *Available at SSRN 6097646*.

Sharma, P. (2026). The Neurocomputational Cost of Cognitive Convenience: How Generative AI Reshapes the Biological Valuation of Effort. *Available at SSRN 7494438*.

Ratajczyk, D., Dymarska, A., Matłoka, A., Tomczyk, M., & Wiącek, M. Thinking With AI, Reasoning Without It: Cognitive Offloading to Generative AI Weakens Effort Regulation.

Atud, V. (2023). *Reclaiming focus in the age of AI: Strategies for deep thinking in a distracted culture*. Sunburst Markets.

Khan, M., Arzu, F., Amjad, S. I., Mahmood, A., & Abrar, K. (2026). Exploring how fintech entrepreneurs build consumer trust in emerging markets amid low financial literacy and high technological skepticism challenges adoption rates. *Bulletin of Management Review*, *3*(2), 524-552.

Guevara, W., & Atomi, C. (2026). The Cognitive Windfall: Where does Human Time, Attention, and Effort Go when AI Lowers the Cost of Cognition?. *Attention, and Effort Go when AI Lowers the Cost of Cognition*.

Klein, C. R., & Klein, R. (2025). The extended hollowed mind: why foundational knowledge is indispensable in the age of AI. *Frontiers in Artificial Intelligence*, *8*, 1719019.

Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Mir, M. M., Mir, S. M., Raina, N. T., Khalid, T. B. A., Nandi, P., Mir, R., ... & Elamin, B. K. (2026). AI Dependency and Cognitive Offloading in Students: Implications for Learning, Thinking, and Competence in the Digital Era. *International Journal of Special Education*, *41*, 30.

Abraham, A. Y., Arthur, F., & Inkoom, M. (2027). Toward Human-Centered AI in Economics Education: Designing Artificial Intelligence for Thinking, Not Cognitive Replacement. In *Psychological and Neurological Consequences of AI: Dependency, Laziness, and Reduced Critical Thinking* (pp. 505-540). IGI Global Scientific Publishing.

Noor, N., Arzu, F., Qadeem, S., Siddique, M., Safi, A., & Qamer, A. (2025). A Big Data–Driven Optimization Framework for Enterprise Financial Management: Enhancing Predictive Decision-Making, Risk Control, and Computational Efficiency. *The Asian Bulletin of Big Data Management*, *5*(4), 190-217.

Riva, G. (2026). The architecture of cognitive amplification: Enhanced cognitive scaffolding as a resolution to the comfort-growth paradox in AI-supported learning. *Cognitive Systems Research*, 101500.

Suryanto, A. F., Wajdi, M., Sahoo, P. K., Wu, Q., Yusof, N. M., & Chan, Y. F. (2025). The impact of artificial intelligence on logic skills, critical thinking, and student creativity. *Journal of Language, Literature, Social and Cultural Studies*, *3*(3), 283-299.

Spector, J. M., & Ma, S. (2019). Inquiry and critical thinking skills for the next generation: from artificial intelligence back to human intelligence. *Smart Learning Environments*, *6*(1), 8.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, *4*, 206.

ASMA, E., SADAF, R., SOMIA, S., FARKHANDA, N., ZAHOOR, A., & NOOR, U. (2025). Influence of Screen Time and Childhood Trauma on Attachment Styles and Emotional Regulation in Adulthood. *JOURNAL FOR SOCIAL SCIENCE ARCHIVES Учредители: Ali Institute of Research & Skills Development*, *3*(1), 312-323.

Rind, I. A. (2026). Conceptualizing the impact of AI on teacher knowledge and expertise: A cognitive load perspective. *Education Sciences*, *16*(1), 57.

Ahmad, B., Muneer, M., Jillani, S. A., & Ahmed, S. (2026, January). Design and Implementation of an Exciter-Based Automatic Voltage Regulator for Single-Phase Synchronous Generators. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Deckker, D., & Sumanasekara, S. (2025). A systematic review of the impact of artificial intelligence, digital technology, and social media on cognitive functions. *International Journal of Research and Innovation in Social Science*, *9*(3), 134-154.

Ejaz, A., Aqeel, A., Latif, N., Ahmed, M. R., & Fatima, F. (2025). Exploring The Effects Of Post-Stroke Cognitive Impairment And Dementia On Patients' Quality Of Life And Independence.

van Zyl, L. E. (2026). The competence paradox: when psychologists overestimate their understanding of artificial intelligence. *Ai & Society*, 1-15.

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain sciences*, *15*(2), 203.

Seung, Y., & Basham, J. D. (2026). Cognitive offloading in the age of generative AI: What does it mean for students with learning disabilities?. *Learning Disability Quarterly*, *49*(3), 121-133.

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y., Nerantzi, C., Moore, S., ... & Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. *Open Praxis*, *16*(4), 487-513.

Aurangzeb, M., & Uddin, S. S. (2025). The artificial intelligence (AI) era: Challenges and opportunities for Pakistan and the Global South. Journal of Current Sign Review, 3(2), 254–264. https://currentsignreview.com/index.php/JCS/article/view/155/115

Zhang, T., Jia, M., & Xue, P. (2025). Impact of AI recommendation styles on carbon management technicians' trust: Dual process of cognitive effort and emotion. *International Journal of Industrial Ergonomics*, *110*, 103822.

Feroz, S., Ahmad, B., Talani, R. A., & Rehman, M. A. U. (2026, March). Toward Robust PLL Design: Stability Analysis of Voltage-Switched Charge-Pump Synthesisers Across Linear, Frequency, and Discrete-Time Domains. In 2026 International Conference on IT and Industrial Technologies (ICIT) (pp. 1-7). IEEE.

Abbas, Z., Zhang, J., & Junaid, M. (2026). When learning meets the machine: How AI-assisted tools reshape deep and surface learning in higher education. *Computers in Human Behavior Reports*, 101057.

Alhur, A. A., Khlaif, Z. N., Hamamra, B., & Hussein, E. (2025). Paradox of AI in higher education: Qualitative inquiry into AI dependency among educators in Palestine. *JMIR Medical Education*, *11*, e74947.

Valcea, S., Hamdani, M. R., & Wang, S. (2024). Exploring the impact of ChatGPT on business school education: Prospects, boundaries, and paradoxes. *Journal of Management Education*, *48*(5), 915-947.

Chan, C. K. Y., & Lee, K. K. (2025). The balancing act between AI and authenticity in assessment: A case study of secondary school students' use of GenAI in reflective writing. *Computers & Education*, *238*, 105399.

Abdi Tabari, M., Rahmati, P., & Huang, Z. (2026). Patterns of learner engagement with generative AI in English writing: cognitive effort and critical thinking. *Innovation in Language Learning and Teaching*, 1-26.

Ning, Y., Sun, S., Lu, Y., Dong, W., Kong, C., & Liu, Y. (2026). Mapping the evidence on generative artificial intelligence for higher-order thinking skills in nursing students: A scoping review. *Nurse Education in Practice*, 104948.

Naz, Z., Baig, H. M., & Parveen, S. (2026). Artificial intelligence and the decline of critical thinking: A hidden academic crisis in contemporary education systems. *Review of Education, Administration & Law*, *9*(1), 53-65.

Rowe, L. (2026). From Gradual Release of Responsibility to Gradual Release of Technology: A Case for the Staged Use of AI in Formal Education. *Education Sciences*, *16*(8), 1291.

Chatto, M., & Logronio, F. (2025). Navigating AI in academia: Undergraduate experiences with ChatGPT and the redefinition of academic writing. *Journal of Education and Learning Reviews*, *2*(4), 27-42.

Demir, M., & Ünveren, D. (2026). The relationship between perceived artificial intelligence competencies and critical thinking skills: The role of digital literacy and cognitive engagement. *Australasian Journal of Educational Technology*, *42*(4), 113-128.

Ozioma, G. N., Obeta, S., Ibanga, D. I., Oraegbunam, L., Itua, R., Amaefule, D. A., ... & Anumaka, C. (2026). AI as cognitive ecology: Revealing the invisible cognitive, cultural, and epistemic costs of generative models. *SSRN Electronic Journal*, 1-16.

Odilova, M., & Nawaz, I. (2025). The Impact of Social Media on Teenagers: Academic Performance, Economic Behaviors, and Mitigation Strategies. *" Navigating Green Economy through Innovations, Policies and Strategies on Sustainable Practices*, 162.

Aurangzeb, M. (2025). Artificial intelligence era: Challenges and opportunities for Pakistan and Global South. https://www.morebooks.de/shop-ui/shop/product/9786209354588

Soni, A., Kumar, M., Kumar, S., Nawaz, I., Mohan, J. M., & Janardhan, R. (2025). The Role of Artificial Intelligence in Transforming Human Resource Management: Opportunities and Challenges in Ethical and Social Issues of Digitalization. *Advances in Consumer Research*, *2*, 1084-1097.



2026 by the authors; Journal of *Kashmir Journal of Academic Research and Development*. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).