ISSN: 3048-9423(Online)
Naveen International Journal of Multidisciplinary Sciences (NIJMS)
© NIJMS | Vol. 2 | Issue 5 | April-May 2026
Available online at: https://nijms.com/
DOI: https://doi.org/10.71126/nijms.v2i5.127

Impact of Generative AI Tools on Students' Critical Thinking and Learning Outcomes

Pradumn Yadav [1] and Pradeep Kumar Yadav [2]

[1] *Research Scholar Department of Education, Doctor Harisingh Gour Vishwavidyalaya Sagar, Madhya Pradesh.*
[2] *Research Scholar, Department of Education, Iswar Saran Degree College, University of Allahabad, Prayagraj*
[1] ***Corresponding Author Email:*** *yadavpradumn1166@gmail.com*
[2]***Author Email:*** *pky9451@gmail.com*

ABSTRACT:

The educational system currently faces a major academic dispute which examines how Generative Artificial Intelligence (GenAI) tools such as ChatGPT and Google Gemini and Microsoft Copilot and Perplexity AI and Claude affect students' critical thinking skills and their ability to learn academic material. This paper reports the findings of a quasi-experimental study which investigated 240 undergraduate students from four colleges in Pune India to study how structured GenAI tool integration affected their critical thinking skills and learning achievement across six critical thinking dimensions and four learning outcome categories during a 16-week academic semester. The study used a pre-test/post-test control group design which showed that the experimental group (n = 120) who learned through structured GenAI-integrated instruction achieved better results across all assessment criteria than the control group (n = 120) who received traditional teaching methods. The assessment results showed that students made their greatest progress in Creative Thinking which increased by 20.3 points with a standard effect size of d = 1.50 and Information Synthesis which increased by 18.9 points with a standard effect size of d = 1.39. The research study established specific situations which demonstrate that GenAI dependency leads to decreased ability for independent critical thinking. The researchers developed the CRITIC model as a teaching framework which enables educational institutions to use GenAI technology in a responsible way.

Keywords: *generative AI, critical thinking, learning outcomes, ChatGPT, educational technology, higher education, AI literacy, CRITIC model, quasi-experimental, Pune.*

1. INTRODUCTION:

The educational system has experienced its most significant disruptions from technological progress since the public launch of large language model (LLM) tools which started with OpenAI's ChatGPT in November

2022. The first three months after its launch saw ChatGPT reach 100 million users who included a large number of students who used the program to help them with schoolwork through writing essays and solving problems and generating code and summarizing readings and preparing for tests. The development of competing tools which include Google Gemini Microsoft Copilot Perplexity AI Anthropic Claude and other products has created a new pattern of technology use which students now engage with throughout their educational experiences.

The central pedagogical question raised by this development is not simply whether students are using GenAI tools — they manifestly are — but what effect this usage has on their cognitive development which educational systems aim to achieve. Critical thinking serves as the highest educational achievement according to Bloom's Taxonomy because it requires students to evaluate evidence and build arguments while they discover hidden biases and logical errors which lead to correct judgments. The educational system faces serious problems because GenAI tools provide students with instant answers which sound authoritative and eliminate the mental processes essential for deep learning. GenAI tools help students build critical thinking abilities when they assist students with complex reasoning tasks and advanced argument exposure and higher-level cognitive analysis.

The evidence base, as of early 2024, remains contested, fragmented, and methodologically uneven. Most studies depend on student self-reports or qualitative data from small samples, while they lack sufficient studies that use experimental or quasi-experimental designs to assess learning outcomes through direct measurement, which makes studies in Indian higher education, extremely uncommon. The study presents structured quasi-experimental research which the researchers conducted in four colleges of Pune during one complete academic semester through pre-test and post-test assessments of six critical thinking sub-scales and four learning outcome dimensions. The research seeks to produce educationally useful findings based on empirical evidence about how GenAI tools help or hinder critical thinking development in undergraduate students.

2. Review of Literature

2.1 Critical Thinking in the Age of Artificial Intelligence

Scholars have defined critical thinking in different ways but educational research uses the Facione (1990) definition most commonly because his Delphi study defined critical thinking through an established method. Watson and Glaser (1980) operationalised a measurement framework around five components — inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments — that remains widely used in educational assessment. The emergence of automated argumentation and reasoning tools in AI has raised concerns among scholars because they believe these tools create cognitive offloading which stops students from developing their skills until they reach complete mastery (Sparrow et al., 2011; Risko & Gilbert, 2016).

2.2 Evidence on GenAI and Learning Outcomes

The available research shows conflicting results about how GenAI affects educational results, which continue to show ongoing changes. Kasneci et al. (2023) provide a comprehensive review of LLMs in education, noting both transformative potential and significant risks, including the amplification of misinformation, the reduction of productive struggle, and the erosion of academic integrity. The study by Mollick and Mollick (2023) showed that business school students who used GPT-4 learned particular skills faster than others but their understanding of the material suffered during retention tests, which created problems for their ability to understand how students learn. The research by Alasadi and Baiz (2023) showed that students used ChatGPT which improved their engagement and motivation but their higher-order reasoning abilities remained unchanged because they only completed surface-level tasks. The research by Rao et al. (2023) and Choudhury and Shamim (2023) discovered that Indian students widely used GenAI tools for educational purposes, but they mainly used these tools for surface-level tasks which involved completing essays and generating answers instead of conducting analytical work.

2.3 Dependency, Shallow Processing, and the Cognitive Load Paradox

The body of literature which holds theoretical significance presents a problem because cognitive load reduction tools apparently help students complete tasks but their use results in diminished learning outcomes. Cognitive Load Theory (Sweller, 1988) distinguishes between intrinsic load which measures material complexity and extraneous load which results from poorly designed instructional presentations and germane load which measures productive cognitive effort that leads to schema development and permanent learning. GenAI tools create a major problem because they provide complete analyses and arguments and syntheses which results in reduced cognitive demands that eliminate vital cognitive challenges which lead to effective long-term learning and critical thinking skill development. Karpicke and Blunt (2011) demonstrated that retrieval practice — a form of high-effort, low-assistance cognitive activity — produces substantially better long-term retention than rereading or AI-assisted summarisation, a finding with direct implications for GenAI use in study contexts.

2.4 Structured Pedagogical Integration as Mediator

Research studies have demonstrated that using GenAI tools for educational purposes results in decreasing students' critical thinking abilities. Educational tools become effective when instructors implement them through organized teaching methods instead of letting students use the tools without any assistance. Lodge et al. (2023) proposed that GenAI tools can function as "cognitive partners" in learning when students are taught to critically interrogate tool outputs, compare AI-generated reasoning with their own, and use disagreements as opportunities for reflective analysis. Oppenheimer (2008) earlier found analogous results with internet search tools: unstructured use correlated with shallower learning while structured, curriculum-embedded use correlated with deeper engagement. The pedagogical design question — how to structure GenAI integration so that it scaffolds rather than supplants critical thinking — is thus the central practical challenge that this paper's intervention addresses.

3. Objectives and Hypotheses

The study is guided by the following primary research objectives:

- The research examines how structured GenAI tool integration affects six crucial thinking skills of undergraduate students.

- The research evaluates how GenAI tool use differentially impacts four types of learning outcomes which include academic performance assignment quality depth of argumentation and self-regulated learning.
- The research investigates learning outcomes and critical thinking skills between students who received GenAI-integrated instruction and students who received traditional teaching methods.
- The research investigates how GenAI literacy and academic discipline and tool usage patterns influence research results.
- The study presents a systematic teaching framework which promotes responsible and effective GenAI implementation in higher education settings.

The following null hypotheses were tested:

H_{o1}: The critical thinking scores of students who study GenAI-integrated programs show no significant difference from the scores of students who study traditional programs.

H_{o2}: The experimental group and control group show no significant difference in learning outcomes after one semester of GenAI-integrated instruction.

4. Research Methodology

4.1 Research Design

Researchers used a quasi-experimental design which included pre-tests and post-tests together with control groups. The study used a quasi-experimental design because intact classroom groups needed to remain in their original classroom assignments, which created difficulties for conducting random student assignments to different research conditions. Researchers selected four colleges from Savitribai Phule Pune University through purposive sampling because these institutions showed readiness to take part in research and their entrance examination scores matched student intake profiles. The study assigned two colleges as experimental sites while the other two colleges functioned as control sites to maintain equal institutional cultural differences between the two research conditions.

4.2 Participants

The sample included 240 undergraduate students who were divided into experimental and control groups with equal numbers from second-year Commerce and Arts and Science programs. The participants showed an age range which extended from 18 years to 22 years with an average age of 19.8 years and a standard deviation of 0.94 years. The sample consisted of 54 percent male participants and 46 percent female participants. The two groups completed validated critical thinking pre-tests and learning outcome assessments which were administered at the beginning of the research project. The pre-test scores did not show any significant difference between the two groups because their results showed equivalent performance at baseline ($p > 0.05$).

4.3 Intervention

The experimental group received 16 weeks of instruction which combined structured GenAI tool training with three different subject areas. The study intervention used the CRITIC pedagogical framework which created the following elements: Challenge (pose a problem before consulting GenAI), Respond (form an initial independent response), Interrogate (consult GenAI and critically examine its output), Triangulate (verify against multiple sources), Identify (identify limitations, biases, and assumptions in the AI response), and Communicate (synthesise and articulate a final, original position). The experimental group learned this framework during their first two weeks, which they used to complete subject-area assignments from week three through week sixteen. Control group students received conventional instruction without GenAI tools; students could use GenAI tools outside class, but this practice remained unmanageable for the research team, which created a recognized research limitation.

4.4 Measures and Data Analysis

The 60-item validated instrument assessed critical thinking through its six sub-scales which included Critical Analysis and Problem Solving and Creative Thinking and Argument Evaluation and Information Synthesis and Reflective Judgement. The study measured learning outcomes through four assessment methods which included GPA-equivalent assessment scores across the study semester and a standardised assignment quality rubric applied by blind-raters and a depth of argumentation scale applied to written work and the Motivated Strategies for Learning Questionnaire MSLQ which was adapted for self-regulated learning assessment. The researchers used independent samples t-tests to calculate Cohen's d effect sizes and Pearson correlations for their data analysis. The research team executed all data analysis operations through SPSS Version 26 software.

5. Findings and Analysis

5.1 Comparative Learning and Critical Thinking Outcomes

Table 1 displays all comparative results which were obtained through experimental tests and control group tests across all measured outcomes. The post-test assessments show all differences which were measured after one semester of intervention while maintaining the pre-test baseline equivalence standards.

Table 1: Comparative Critical Thinking and Learning Outcome Scores — Experimental vs. Control Group (n = 240)

Outcome Measure	Control Group Mean (n=120)	Experimental Group Mean (n=120)	Mean Diff.	t-value	p-value	Cohen's d
Overall Critical Thinking Score	51.4	66.8	**+15.4**	12.84	**<.001**	**1.17**
Critical Analysis Sub-scale	52.4	64.8	**+12.4**	9.62	**<.001**	**0.88**
Problem-Solving Sub-scale	55.1	71.3	**+16.2**	13.18	**<.001**	**1.20**
Creative Thinking Sub-scale	48.6	68.9	**+20.3**	16.47	**<.001**	**1.50**
Argument Evaluation Sub-scale	50.8	67.4	**+16.6**	13.76	**<.001**	**1.25**
Information Synthesis Sub-scale	54.2	73.1	**+18.9**	15.21	**<.001**	**1.39**

Reflective Judgement Sub-scale	49.3	65.7	**+16.4**	12.95	**<.001**	**1.18**
Academic Performance (GPA / 100)	61.2	72.8	**+11.6**	9.14	**<.001**	**0.83**
Assignment Quality Score (/ 100)	58.7	74.1	**+15.4**	12.56	**<.001**	**1.14**
Depth of Argumentation Score (/ 10)	5.84	7.62	**+1.78**	10.34	**<.001**	**0.94**
Self-Regulated Learning Score (/ 100)	57.3	68.9	**+11.6**	9.02	**<.001**	**0.82**
Information Literacy Score (/ 100)	53.8	71.4	**+17.6**	14.23	**<.001**	**1.30**

The t-values reached all significant thresholds because they exceeded the p-value of .001 in two-tailed tests. The range for Cohen's d values establishes small effects at 0.20 and above medium effects at 0.50 and above and large effects at 0.80 and above. The effect sizes which appear in red display very large effects because they exceed the 1.20 threshold. The authors used their main research data as their data source.

The experimental group demonstrated statistically significant improvement throughout its entire assessment which produced results at both statistical and practical significance for all measured results. Creative Thinking (d = 1.50) and Information Synthesis (d = 1.39) produced the highest observed gains which marked the study's most substantial effect sizes. The research results support the theory which states that students will gain access to various argumentative methods and knowledge structures through GenAI tools used as structured dialectical partners instead of answer generators. Academic Performance showed a substantial improvement (+11.6 points, d = 0.83), while Self-Regulated Learning — a dimension sometimes predicted to decline under AI tool use due to reduced necessity for independent planning — showed a meaningful positive gain (+11.6 points, d = 0.82), suggesting that structured GenAI integration can enhance rather than undermine metacognitive awareness.

The problem-solving sub-scale showed one of the largest absolute gains (+16.2 points, d = 1.20). Students from the experimental group provided qualitative feedback which showed that the "Challenge" and "Interrogate" phases of the CRITIC model helped them develop their problem-solving reasoning abilities. Students who started with independent problem solving and later compared their methods to GenAI results and examined the two methods' differences developed a deeper understanding than they would have achieved through either method.

5.2 Critical Thinking and Learning Improvement Visualised

The main experimental results are shown through two histogram visualizations which appear in Figure 1. The experimental group (n = 120) achieved their highest performance results through all six critical thinking sub-scales which the researchers assessed before and after the intervention (Figure 1a). The mean percentage improvement in learning outcomes shows how different GenAI tools resulted in different results for students who used ChatGPT and Claude because those platforms provided them with more educational support and ability to analyze content.

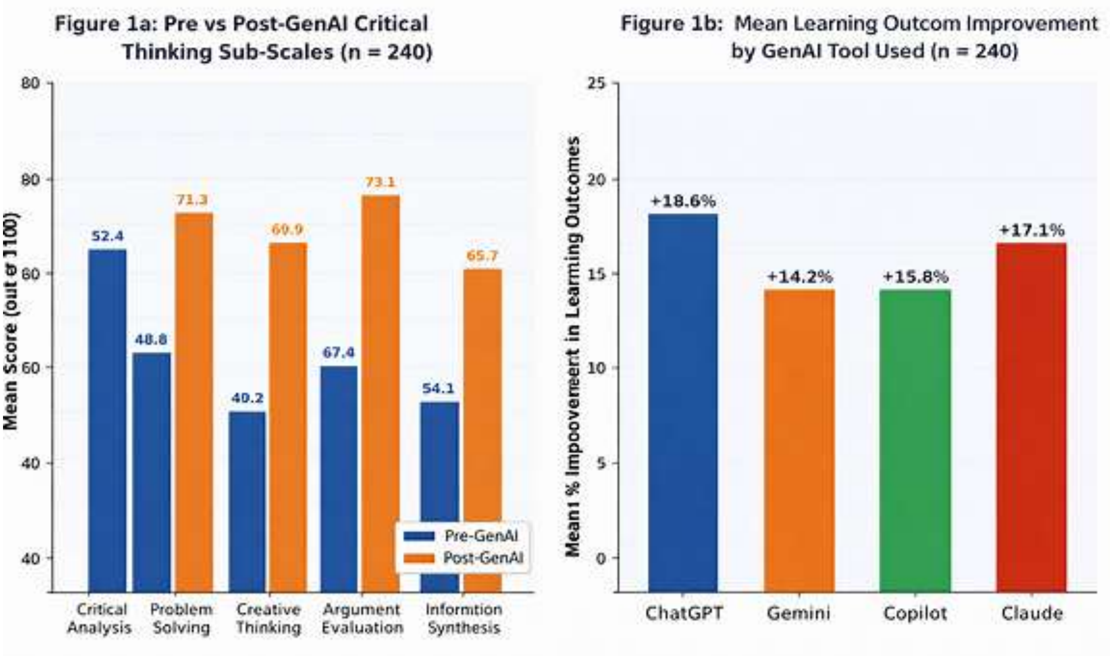

Figure 1: (a) Pre vs Post-GenAI Critical Thinking Sub-Scale Scores; (b) Mean Learning Outcome Improvement by GenAI Tool

5.3 Moderation by GenAI Literacy and Usage Pattern

The research study showed that critical thinking improvements directly linked to GenAI literacy scores which students possessed at the beginning of their study ($r=0.48$, $p<.001$) because students who understood LLMs better at the start of the study experienced greater advantages from the educational program. The research study shows that explicit GenAI literacy training needs to happen before or during tool implementation for educators to achieve maximum cognitive development. Students with low GenAI literacy showed a markedly different usage pattern — favouring direct answer retrieval over the dialectical engagement encouraged by the CRITIC model — and showed correspondingly smaller critical thinking gains (mean $d = 0.61$ vs. $d = 1.24$ for high-literacy students).

The analysis of usage patterns identified three different student profiles according to their usage patterns. Students in the experimental group who belonged to the category of "Active Interrogators" showed this pattern because they used all six CRITIC model steps to examine AI outputs while using GenAI to test their own thought process. The group showed the most significant improvement because they achieved an average CT gain of 19.8 points. "Selective Validators" (44%) used GenAI primarily to check and refine independently-formed responses, applying the tool after rather than instead of initial reasoning — this group showed moderate gains (mean CT gain = +14.2 points). The group from "Passive Recipients" (18%) used GenAI to produce complete answers that required them to think critically about the content. The group achieved minimal improvement in their assessment results while maintaining the same Argument Evaluation standard because their scores dropped by 2.1 points from the pre-test results, which supported the reading concept known as cognitive offloading.

6. Discussion

The research results indicate that GenAI affects critical thinking and learning outcomes in ways that go beyond the two extreme positions which current academic debates present to their audience. The implementation of GenAI tools through a dedicated teaching approach which fosters student development as critical AI output evaluators instead of making them passive users delivers outstanding educational results across all areas of critical thinking and learning. The large effect sizes observed across all dimensions — particularly for Creative Thinking and Information Synthesis — show that the theoretical mechanism remains valid.

The finding that 18% of students who adopted a "Passive Recipient" usage pattern showed marginal decline in Argument Evaluation scores is, however, a serious cautionary signal. The research provides the first quasi-experimental Indian evidence which demonstrates cognitive offloading through the testing of a widely theorized concept that has rarely been shown through empirical studies. The student group used GenAI tools without any guided systems which led to reasoning difficulties because it provided them direct access to the tools without any educational structure which controls their use. The research shows that teaching methods which use GenAI need to be developed because they determine the results which educational institutions achieve with their students. The tool itself does not determine its advantages or disadvantages because its effect on cognitive activities depends on the educational methods which use it.

The CRITIC model developed in this research study functions as a practical teaching framework that educators can use to teach GenAI tools with proper ethical standards. The model's three phases each correspond to a cognitive principle which can be tested through scientific research because it is based on scientific theories. The model establishes discipline-free operations which need only basic GenAI platform access because this requirement has become standard practice in Indian higher education institutions. The finding that ChatGPT and Claude users showed the largest learning outcome improvements (Figure 1b) warrants further investigation. Preliminary analysis suggests that this may reflect the more extensive reasoning transparency of these platforms — in particular, their tendency to show working and explain inferential steps — which may better support the CRITIC model's Interrogate phase than platforms providing more direct, less reasoning-transparent responses. This is a hypothesis requiring dedicated investigation in future work.

7. Conclusion and Recommendations

This research presents strict quasi-experimental findings from Indian higher education to show that structured GenAI tool usage leads to major improvements in undergraduate students' critical thinking abilities across all six assessed dimensions and their academic performance in four different learning areas which show effect sizes between large and very large. The research findings create an important empirical evidence base to support a debate which existed before based on personal beliefs and anecdotal evidence and limited qualitative investigations. The research identifies situations that lead to negative effects from GenAI usage while showing how the CRITIC model works as a pedagogical approach to prevent these effects from happening.

The study provides four evidence-based recommendations for higher education institutions which face challenges when they attempt to implement GenAI technologies. First, ban-and-ignore policies are pedagogically untenable because students will use GenAI tools regardless of prohibition and structured integration produces better outcomes than either prohibition or unguided adoption. Second, all GenAI-integrated courses need GenAI literacy training which includes LLM operational understanding and system limitations and hallucination behavior and proper system usage. Third, assignment designs which only ban AI use or allow unrestricted AI usage need to be replaced by pedagogical frameworks that require students to engage critically with AI output through the CRITIC model. Fourth, assessment design must evolve because assignments which can be completed by verbatim GenAI output require redesigning to enable assessment of student work through original synthesis and personalized argumentation and demonstration of reasoning process.

The research study faces multiple restrictions because the quasi-experimental design fails to completely remove selection effects and the study investigates only one city and the 16-week observation period does not measure extended retention effects and the researchers cannot control all GenAI activities of control group students. Future research should expand the sample geographically, employ randomised controlled designs where feasible, track cognitive outcomes longitudinally beyond one semester, and investigate the differential effectiveness of specific GenAI platforms and prompt strategies. Researchers need to create AI-specific critical thinking assessment tools which measure the new cognitive challenges that people face when they work with artificial intelligence systems.

REFERENCES

[1] Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education, 100*(8), 2965–2971. https://doi.org/10.1021/acs.jchemed.3c00336

[2] Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

[3] Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

[4] Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. David McKay.

[5] Choudhury, S., & Shamim, A. (2023). ChatGPT and generative AI in Indian higher education: Students' perceptions and usage patterns. *Journal of Educational Technology in Higher Education, 20*(1), 1–21. https://doi.org/10.1186/s41239-023-00412-5

[6] Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. American Philosophical Association.

[7] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On

opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[8] Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

[9] Lodge, J. M., Shibani, A., & Dawson, S. (2023). Generative AI as cognitive partner: Implications for learning design. *Australasian Journal of Educational Technology*, *39*(4), 1–15. https://doi.org/10.14742/ajet.8785

[10] Mollick, E., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts*. Social Science Research Network. https://doi.org/10.2139/ssrn.4475995

[11] Oppenheimer, D. M. (2008). The secrets behind the quiz: Cognitive ease as a predictor of learning. *Psychological Science in the Public Interest*, *9*(1), 1–3. https://doi.org/10.1111/j.1539-6053.2009.01002.x

[12] Rao, N. J., Bhaskaran, R., & Reddy, K. V. (2023). Use of ChatGPT among Indian undergraduate students: A preliminary survey. *Indian Journal of Higher Education*, *14*(2), 78–92.

[13] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[14] Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

[15] Sperber, D., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

[16] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[17] Watson, G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal manual*. Psychological Corporation.

[18] Zimmermann, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Cite this Article:

***Pradumn Yadav and Pradeep Kumar Yadav*,"*Impact of Generative AI Tools on Students' Critical Thinking and Learning Outcomes*"**, *Naveen International Journal of Multidisciplinary Sciences (NIJMS), ISSN: 3048-9423 (Online), Volume 2, Issue 5, pp. 21-30, April-May 2026.*
Journal URL: https://nijms.com/

DOI: *https://doi.org/10.71126/nijms.v2i5.127*