Journal of
Education and Practice
(JEP)

# Guided Integration or Prohibition? A Quasi-Experimental Study of Generative AI in Project-Based Learning

CARI Journals

Journal of Education and Practice
ISSN 2520-467X (Online)
Vol. 10, Issue No. 6, pp. 59 - 74, 2026
www.carijournals.org

# Guided Integration or Prohibition? A Quasi-Experimental Study of Generative AI in Project-Based Learning

**Marine Aghekyan[1*], Marie Botkin[2]**

[1,2]California State University

Long Beach, 1250 Bellflower Blvd, Long Beach, CA 90840, USA

https://orcid.org/0009-0009-6347-0847

***Accepted: 25th June, 2026, Received in Revised Form: 2nd July, 2026, Published: 21st July, 2026***

## Abstract

**Purpose:** The rapid expansion of generative artificial intelligence (AI) tools in higher education has prompted institutional responses ranging from prohibition to structured integration. However, empirical evidence comparing these approaches within authentic classroom contexts remains limited. This quasi-experimental study examines whether guided AI integration enhances student learning outcomes in a project-based undergraduate course.

**Methodology:** Two intact sections of a senior-level global trade course (N = 60) completed an identical semester-long group project. One section operated under a structured AI policy (Experimental Group), while the other followed a strict human-only policy (Control Group).

**Findings:** Independent-samples t-tests revealed statistically significant differences in overall project performance favoring the Experimental Group across written and presentation components. Students in the Experimental Group demonstrated stronger analytical organization, clearer conceptual application, and more effective synthesis of course content. Self-reported measures further indicated higher perceived learning, greater professional growth, and reduced stress. Indicators consistent with AI-assisted writing were observed in some Control Group submissions; however, these observations were not formally quantified and are interpreted cautiously.

**Unique Contribution to Theory, Policy and Practice:** Findings suggest that structured and transparent AI integration, when aligned with instructional design principles, may support stronger performance outcomes than prohibition-based approaches within this course context. The study contributes empirical evidence to ongoing discussions regarding AI governance in higher education and offers implications for instructional policy design.

**Keywords:** *Generative AI, Project-Based Learning, Higher Education, Instructional Policy, Academic Integrity*

## INTRODUCTION

The rapid emergence of generative artificial intelligence (AI) tools such as ChatGPT has introduced significant disruption to higher education. Within months of their public release, students began integrating these tools into research, writing, and collaborative coursework, often in the absence of clear institutional guidance. Universities responded unevenly. Some adopted prohibition-based policies grounded in academic integrity concerns, while others experimented with structured integration frameworks. Despite widespread debate, empirical evidence directly comparing these approaches within authentic instructional contexts remains limited.

Emerging scholarship suggests that the educational impact of generative AI depends less on the technology itself and more on how it is pedagogically structured. Systematic reviews and meta-analyses indicate that AI-assisted learning may produce small-to-moderate gains when aligned with instructional objectives (Deng et al., 2025; Naznin et al., 2025). Conversely, unstructured or unguided use may reduce cognitive engagement and lead to superficial processing (Lo, 2023). These findings highlight a central tension: whether AI functions as a cognitive scaffold or a shortcut appears to depend primarily on instructional design rather than technological capability alone.

Project-based learning (PBL) provides a particularly relevant environment for examining this issue. PBL requires sustained inquiry, analytical synthesis, collaborative problem-solving, and iterative revision. These processes demand higher-order reasoning and metacognitive engagement, conditions under which generative AI tools may either support or displace meaningful learning. Because project work unfolds over extended periods, it offers a meaningful context for evaluating how policy framing influences AI use and learning outcomes.

Although institutions increasingly issue AI-related policies, few empirical studies directly compare structured AI integration with formal prohibition within the same course environment. Much of the existing research relies on short-term laboratory tasks or isolated writing assignments rather than semester-long, collaborative projects (Bond et al., 2024). As a result, higher education institutions lack classroom-based evidence to inform AI governance decisions.

This study addresses that gap by implementing a quasi-experimental comparison between two sections of the same undergraduate course operating under contrasting AI policies: one permitting structured and transparent AI use (Experimental Group) and one enforcing a human-only policy (Control Group). By examining project performance, perceived learning, confidence, and academic integrity perceptions, this research contributes empirical evidence to inform ongoing debates about AI governance in higher education.

The study addresses four research questions:

1. Does guided generative AI integration improve overall project performance compared to a human-only condition?

2. Does guided AI integration influence students' perceived learning, confidence, and satisfaction?
3. How does instructional AI policy shape students' academic integrity perceptions and ethical awareness?
4. What differences emerge in student-reported experiences between structured and prohibited AI conditions?

Students are already using AI, sometimes without transparency, sometimes without understanding its limits and its opportunities. Instructors and institutions need clear understanding to guide decisions. This research study aims to provide that evidence by analyzing outcomes within a single course structure, using senior level student work, survey results, and policy-based conditions. The goal is not to promote or discourage AI use, but to understand the pedagogical consequences of structured, transparent integration compared to prohibition.

## LITERATURE REVIEW

### Generative AI and Learning Outcomes in Higher Education

The rapid integration of generative artificial intelligence (AI) tools into higher education has prompted growing empirical investigation into their influence on student learning. Early research on artificial intelligence in education largely focused on intelligent tutoring systems, predictive analytics, and automated assessment (Zawacki-Richter et al., 2019). However, the emergence of large language models capable of generating coherent academic text has shifted scholarly attention toward questions of authorship, cognitive engagement, and instructional effectiveness.

Systematic reviews examining the educational impact of generative AI report mixed but cautiously optimistic findings. Lo (2023), in one of the earliest rapid reviews of ChatGPT in education, noted potential benefits for drafting support, idea generation, and formative feedback, while simultaneously cautioning against overreliance. More recently, Deng et al. (2025) conducted a meta-analysis of experimental studies investigating ChatGPT-assisted learning and found small-to-moderate positive effects on academic performance, particularly in writing-intensive tasks. However, the authors emphasized that learning gains were highly contingent on instructional framing and task design.

Similarly, Naznin et al. (2025), in a systematic review of generative AI integration in higher education, concluded that AI tools can enhance writing fluency, brainstorming efficiency, and coding performance when embedded within structured pedagogical frameworks. The review stressed alignment between AI use and learning objectives as a critical determinant of effectiveness. In contrast, unguided or unrestricted AI use was associated with surface-level engagement and diminished knowledge retention.

Taken together, existing evidence suggests that the educational impact of generative AI is not inherent to the technology itself but mediated by pedagogical design. When AI use is embedded within explicit instructional frameworks and aligned with learning goals, it may support higher-order reasoning. When used without guidance, it may contribute to superficial engagement. However, much of the current literature relies on short-term tasks or isolated assignments rather than semester-long coursework. These contrasting findings suggest that the educational impact of generative AI is not inherent to the technology itself but mediated by pedagogical design. The central empirical question emerging from this literature is whether AI functions as a cognitive amplifier or as a cognitive shortcut. When positioned as a tool for structuring thought, prompting reflection, or supporting revision, AI may enhance higher-order reasoning. When used to bypass effortful thinking, it may undermine learning processes.

Despite the growing body of research, important gaps remain. First, many studies rely on short-term laboratory tasks or controlled experimental exercises rather than semester-long coursework. Second, much of the empirical evidence focuses on individual performance rather than collaborative, project-based contexts. Third, few studies directly compare structured AI integration with formal prohibition within the same instructional environment. As Bond et al. (2024) observe in their meta-systematic review, AI-in-education research requires more contextually grounded, methodologically rigorous studies capable of informing real-world policy decisions.

The present study responds to these gaps by examining generative AI integration within a semester-long, collaborative project-based course. By comparing a structured AI-guided condition with a human-only policy condition, this research moves beyond isolated task performance to evaluate sustained project outcomes, perceived learning, and academic integrity perceptions within an authentic instructional setting. In doing so, it contributes empirical evidence to clarify how instructional framing shapes the relationship between generative AI use and student learning outcomes.

**Instructional Design, Cognitive Effort, and Project-Based Learning**

The variability in reported outcomes of generative AI use in education suggests that instructional design, rather than technology alone, determines whether AI enhances or undermines learning. From a cognitive perspective, meaningful learning requires effortful processing, elaboration, and metacognitive monitoring. When students engage deeply with content, organizing information, evaluating sources, generating explanations, they are more likely to achieve durable understanding. Conversely, when tools reduce cognitive effort without supporting conceptual integration, learning may become superficial.

Cognitive Load Theory (Sweller, 1988) provides a useful framework for understanding this dynamic. Learning is constrained by the limited capacity of working memory. Effective instructional design reduces unnecessary cognitive load while preserving germane load associated with schema construction. Generative AI tools may reduce extraneous load by assisting with

information organization or initial drafting; however, they may also reduce germane load if students outsource core reasoning tasks. Thus, whether AI serves as a cognitive scaffold or a cognitive shortcut depends on how it is integrated into learning activities.

Kirschner, Sweller, and Clark (2006) argue that minimally guided instruction often fails because novices require structured support to manage complex problem-solving tasks. In this context, generative AI can be conceptualized as a form of guided assistance, if appropriately framed. When instructors provide explicit expectations, verification requirements, and transparency guidelines, AI may function as an adaptive scaffold rather than an unregulated shortcut. The distinction lies not in the presence of AI, but in the degree of instructional oversight.

The concept of scaffolding further clarifies this distinction. Wood, Bruner, and Ross (1976) describe scaffolding as temporary support that enables learners to perform tasks beyond their independent capacity while gradually transferring responsibility back to the learner. If generative AI is positioned as a scaffold, supporting brainstorming, outlining, or revision while requiring human evaluation and synthesis, it may extend cognitive capacity without replacing intellectual effort. However, when AI generates completed work with minimal student engagement, scaffolding collapses into substitution.

Self-regulated learning theory adds another layer to this analysis. Zimmerman (2002) emphasizes that effective learners actively plan, monitor, and evaluate their cognitive processes. AI integration may either support or disrupt self-regulation. Structured AI policies that require documentation, verification, and reflection may encourage metacognitive engagement. In contrast, unrestricted AI use may weaken monitoring processes, allowing students to accept generated content without critical evaluation.

Project-based learning (PBL) offers a particularly relevant environment for examining these interactions. PBL is characterized by sustained inquiry, collaborative problem-solving, authentic tasks, and iterative refinement (Hmelo-Silver, 2004). Students must integrate multiple sources, evaluate trade-offs, justify decisions, and communicate coherent arguments. These tasks require higher-order cognitive processes aligned with analysis, synthesis, and evaluation. Because PBL unfolds over extended periods and demands coordination among team members, it creates both opportunities and risks for AI integration. Generative AI may support early-stage ideation, information structuring, or draft refinement, potentially increasing efficiency and reducing extraneous cognitive burden. At the same time, overreliance on AI-generated text may dilute collaborative negotiation, reduce deep engagement with primary sources, and weaken conceptual ownership.

Despite the theoretical relevance of PBL as a testing ground for AI integration, empirical research directly examining generative AI within semester-long project-based contexts remains limited. Much of the current literature relies on short-term experiments or individual writing tasks. There is comparatively little evidence evaluating how AI influences sustained collaborative projects, iterative revision processes, and multi-dimensional performance assessments. The

present study addresses this gap by embedding generative AI within a semester-long project-based course under structured instructional guidelines. Rather than treating AI as a standalone tool, the study positions AI use within a clearly articulated policy framework emphasizing transparency, verification, and human oversight. By comparing this structured integration model with a human-only condition in the same course context, the study examines whether AI reduces harmful cognitive shortcuts or instead supports scaffolded engagement in complex project-based work.

Through this design, the research contributes to a broader theoretical question: under what instructional conditions does generative AI function as a cognitive amplifier rather than a cognitive substitute? By situating AI integration within cognitive load theory, supporting theory, and self-regulated learning frameworks, the study moves beyond technological novelty and toward a principled analysis of instructional design.

**Institutional Policy, Academic Integrity, and AI Governance**

The rapid emergence of generative AI has outpaced institutional policy development. Universities have responded with varied approaches, including formal bans, partial restrictions, or permissive guidelines emphasizing disclosure. Policy ambiguity has contributed to uncertainty among faculty and students. Academic integrity concerns remain central to institutional debates. Balalle and Pannilage (2025) note that AI-generated text challenges traditional definitions of authorship and originality. Cotton et al. (2023) highlight that concerns about misuse often drive prohibition-based policies. Holmes et al. (2022) argue that ethical AI integration requires explicit instruction in transparency, verification, and responsible oversight rather than reliance solely on detection mechanisms.

Bond et al. (2024) further emphasize the need for rigorous, classroom-based comparative research to guide policy decisions. In particular, few studies directly compare structured AI integration with formal prohibition within the same course context. As a result, institutional AI governance decisions are frequently made in the absence of controlled classroom-level evidence.

In response to these gaps, the present study implements a quasi-experimental comparison of two intact course sections operating under contrasting AI policies: a structured integration model and a human-only prohibition model. By examining project performance, perceived learning outcomes, and student experiences within a semester-long project-based course, this study contributes empirical evidence to inform ongoing discussions about AI governance in higher education.

## METHOD

This study employed a quasi-experimental, non-equivalent group design comparing two intact sections of an upper-division, project-based undergraduate course. Random assignment was not feasible because students enrolled in course sections prior to the implementation of instructional AI policies. The two sections were therefore treated as naturally occurring comparison groups. Both sections received identical instructional content, assignments, grading

rubrics, timelines, and assessment criteria. The sole systematic difference between sections was the instructional policy governing the use of generative AI tools during completion of the semester-long project. One section operated under a structured AI-integration policy (Experimental Group) that permitted the use of generative AI tools within clearly defined guidelines emphasizing transparency, verification, and human oversight. The second section followed a human-only policy prohibiting AI use (Control Group) at any stage of the project. This design enabled examination of student outcomes under two contrasting instructional policy conditions within the same course context.

Participants were 60 undergraduate students enrolled in two sections of a senior-level course in global trade and sourcing at a large public university. Each section included 30 students. The majority of participants were juniors or seniors majoring in fashion merchandising or fashion design. The course centers on a comprehensive, semester-long group project in which students develop an outsourcing strategy for a U.S.-based fashion company expanding production internationally. Project requirements include vendor evaluation, country risk assessment, cost analysis, sustainability considerations, and development of a strategic sourcing plan. Students worked in self-selected teams of four to five members. Participation in the research component was voluntary. Students reviewed and signed informed consent forms during the first two weeks of the semester. No student declined participation. The study was reviewed and approved under institutional guidelines for research on instructional practices. Importantly, aside from the AI policy condition, all instructional activities, evaluation procedures, and performance expectations were equivalent across sections.

**Experimental Group (Structured AI Use)**

Students in Experimental Group were permitted to use generative AI tools as structured learning support during completion of the semester-long project. The instructional policy explicitly defined acceptable and prohibited uses of AI. Permitted uses included idea generation, organizational outline, clarification of concepts, drafting assistance, and scenario exploration. Prohibited uses included submission of fully AI-generated work without revision, fabrication of sources or data, bypassing analytical reasoning, or presenting unverified AI output as final analysis. At the beginning of the project phase, students in this section participated in an instructor-led orientation outlining responsible AI use expectations. Students were required to sign an AI policy acknowledgment form and maintain an AI use log documenting (a) the tool used, (b) the purpose of use, and (c) evidence of human review and revision. All AI-assisted content incorporated into the final project required independent verification through credible external sources. Students were also required to disclose AI use in an appendix to their written report. Instructor check-ins during project development included reminders regarding fact verification, ethical standards, and alignment between AI-assisted drafts and course learning objectives. The instructional emphasis positioned AI as a support tool rather than a substitute for student reasoning.

**Control Group (AI Prohibition)**

Students in the Control Group completed the identical semester-long project under a human-only policy. The instructional policy prohibited the use of generative AI tools at any stage of the project, including brainstorming, drafting, paraphrasing, summarizing, translation, or use of embedded AI features within commonly used writing software. Students in this section signed a human-only policy acknowledgment form affirming that all submitted work would be independently produced without AI assistance. Instructor check-ins emphasized traditional research processes, source evaluation, drafting, and revision strategies. Faculty reserved the right to request outlines, early drafts, or explanations of analytical decisions to verify compliance with course expectations. Aside from the AI policy condition, all instructional materials, deadlines, grading rubrics, feedback procedures, and evaluation criteria were identical across sections.

**Procedures**

Weeks 1-2: Consent and pre-test administration. Students completed a Pre-Test Survey measuring baseline knowledge, project confidence, AI attitudes, and academic integrity perceptions. Weeks 3-6: Project launch and policy implementation. The Experimental Group received AI orientation; the Control Group reviewed academic integrity guidelines. Weeks 7-12: Project development. Teams conducted research, drafted written reports, and participated in structured in-class activities. The Experimental Group used AI within structured guidelines; the Control Group used traditional methods. Weeks 13-15: Oral presentations were delivered and evaluated using a standardized rubric. Students completed the Post-Test Survey following presentations.

Beginning in Week 6 of the semester, both sections participated in structured, in-class group activities designed to develop skills required for the final project. Part of the final project required students to evaluate and select a vendor based on multiple criteria, including production capacity, sustainability practices, and human-rights considerations. Following instruction on the vendor selection chapter, students were assigned a randomly selected retailer and tasked with identifying an appropriate outsourcing vendor that aligned with the retailer's mission, size, brand image, and market positioning. Students in the Experimental Group were permitted to use an AI tool as an initial exploratory aid, limited to a single prompt. After generating AI-assisted output, students were required to conduct independent internet-based research to verify information, evaluate source credibility, and refine their analyses. The instructor provided formative feedback during the activity, emphasizing critical engagement with AI-generated content and alignment with academic integrity standards.

Students in the Control Group completed the same activity using traditional internet-based research methods only and were not permitted to use AI tools. Instructional guidance and instructor feedback during the activity focused on source evaluation, evidence-based reasoning, and alignment with course objectives, mirroring the feedback provided to the experimental section aside from AI-related guidance. Similar in-class activities were conducted focusing other portions

of the final project including outsourcing country assessment, payment methods, shipping, negotiating, etc.

**Measures**

*Project Performance*

Student performance was assessed using a standardized grading rubric applied consistently across both course sections. The rubric evaluated written reports and oral presentations on criteria including analytical depth, conceptual accuracy, organization, quality of evidence, creativity, and delivery effectiveness. The total project score was calculated out of 100 points, combining written and presentation components. All grading criteria were identical across instructional conditions.

**Pre-Test and Post-Test Surveys**

*Pre-test Survey*

At the beginning of the semester, students completed a pre-test survey designed to assess baseline knowledge, confidence, AI attitudes, and ethical awareness prior to implementation of the AI policy conditions. All Likert-scale items were measured on a 5-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The pre-test included the following constructs: global sourcing knowledge (4 items; items assessed students' perceived understanding of key global sourcing concepts, sustainability considerations, ethical supply chain issues, and cost/risk factors); project development confidence (4 items; items measured confidence in managing group projects, organizing research tasks, collaborating effectively, and critically evaluating sources); attitudes toward AI in learning (4 items; items assessed beliefs regarding AI's usefulness for brainstorming and writing, concerns about reduced critical thinking, and expectations regarding AI disclosure); ethical awareness and academic integrity (4 items; items evaluated understanding of institutional academic integrity policies and perceptions of ethical versus unethical AI use). In addition, students reported prior experience with generative AI tools (frequency and type of use), providing contextual baseline data. Composite scale scores were calculated for each construct following reliability analysis.

*Post-Test Survey*

Following completion of the project and presentations, students completed a post-project survey assessing perceived learning outcomes, confidence, collaboration experiences, and AI-related reflections. All Likert-scale items used the same 5-point response format as the pre-test. The post-test included the following constructs: perceived global sourcing knowledge (4 items; items assessed students' perceived understanding of key global sourcing concepts, sustainability considerations, ethical supply chain issues, and cost/risk factors understanding of sourcing strategies, industry applications, and key challenges); project confidence and collaboration (3 items; measured confidence in project planning, teamwork effectiveness, and task management); attitudes toward AI in learning (3 items; examined enthusiasm toward AI, perceived learning

benefits, and concerns about skill development); ethical awareness (3 items; measured clarity regarding academic integrity policies and perceptions of AI-related misconduct); presentation self-efficacy (4 items; assessed confidence in public speaking, communication clarity, visual design, and audience engagement); perceived professional growth (4 items; measured perceived value of the presentation experience for communication development and future professional readiness); team collaboration and problem-solving (4 items; evaluated perceived effectiveness of group collaboration during presentation preparation). For students in the Experimental Group additional open-ended questions captured reflections on AI use, verification practices, and perceived influence on presentation and writing development.

## RESULTS

Descriptive statistics (means and standard deviations) were calculated for each subscale by instructional condition (Experimental Group vs. Control Group). Distributional characteristics were examined using histograms and boxplots to identify potential outliers or violations of normality assumptions. Independent samples t-tests were conducted on pre-test composite measures to evaluate baseline equivalence between groups prior to implementation of instructional AI policies. Where pre-test data were available for corresponding constructs (e.g., knowledge, confidence, ethical awareness), these measures were used to assess comparability across sections.

To examine differences between instructional conditions, independent samples t-tests were conducted on total project performance scores, written and presentation component scores, and post-test composite survey measures. Effect sizes were calculated using Cohen's *d* to estimate the magnitude of group differences. Ninety-five percent confidence intervals were reported where appropriate to provide additional precision regarding effect estimates. Assumptions of normality and homogeneity of variance were assessed prior to conducting parametric tests. Levene's test was used to evaluate equality of variances.

Open-ended responses regarding AI prompt use, verification practices, and reflections on AI's role in presentation preparation were analyzed using thematic analysis. Responses were reviewed iteratively to identify recurring themes related to brainstorming, organization, revision, verification strategies, and perceived influence on presentation quality. Frequency analysis was conducted on binary verification responses (Yes/No) to assess the proportion of students reporting independent fact-checking of AI-generated content. The study was approved under IRB Exempt Categories 1 and 2, permitting evaluation of instructional techniques and anonymous survey research. Surveys were collected anonymously. Students were reminded that participation or nonparticipation was voluntary. AI use data were disclosed only through self-reported logs and documentation submitted voluntarily.

Descriptive statistics were calculated for all composite pre- and post-test measures. Reliability analyses indicated acceptable internal consistency across multi-item constructs ($\alpha$ values $\geq .80$). Assumptions of normality and homogeneity of variance were examined prior to inferential testing. Visual inspection of distributions and Levene's tests indicated no substantial

violations; therefore, independent-samples t-tests were conducted. Baseline equivalence between sections was evaluated using independent-samples t-tests on all pre-test composites (knowledge, project confidence, attitudes toward AI, and ethical awareness). No statistically significant differences emerged at pre-test (all $p > .05$), indicating that the Experimental Group and Control Group were comparable prior to implementation of the instructional policy conditions.

To test Research Question 1: *Does guided generative AI integration improve overall project performance compared to a human-only condition*, independent-samples t-tests revealed statistically significant differences in overall project performance favoring the Experimental Group. The Experimental Group (M = 89.16, SD = 4.26) scored significantly higher than the Control Group (M = 77.27, SD = 3.23) on total project performance (maximum = 100 points), $t(58) = 12.32$, $p < .001$, $d = 3.08$. The difference was statistically significant, and the effect size was large. When components were analyzed separately, written report performance was significantly higher in the Experimental Group (M = 44.97, SD = 1.94) than in the Control Group (M = 40.29, SD = 0.90), $t(44) = 12.16$, $p < .001$, $d = 3.08$. The difference was statistically significant, and the effect size was large. Oral presentation performance also favored the Experimental Group (M = 44.19, SD = 3.35) compared to the Control Group (M = 42.27, SD = 3.23), $t(58) = 2.29$, $p = .026$, $d = 0.59$, indicating a statistically significant difference with a moderate effect size.

Grade distributions further illustrate the magnitude of these differences. In the Experimental Group, the majority of students scored in the A range (85–100 points), with a tight clustering of scores between 85 and 95. No failing or near-failing projects were observed. In contrast, the Control Group's scores clustered primarily in the C to low B range (70–82 points), with substantially fewer A-level performances. The Experimental Group demonstrated both higher mean performance and reduced variability around high performance levels. The Control Group demonstrated lower overall means and greater evidence of conceptual inconsistency and superficial analysis. Instructor review of written submissions identified indicators consistent with AI-assisted writing in several Control Group submissions, despite signed agreements prohibiting AI assistance. Indicators included repeated use of long-dash formatting characteristics of generative AI outputs, unusual patterns of bolded or italicized emphasis inconsistent with typical student writing, generic phrasing and template-like structural repetition, and occasional factual inconsistencies. These observations were not formally quantified through automated detection tools and are therefore interpreted cautiously.

It is important to note that this unstructured AI engagement did not correspond to improved performance. Control Group projects were frequently characterized by fragmented argumentation, superficial analysis, weaker alignment with project objectives, and limited integration of course concepts. This pattern suggests that prohibition did not eliminate AI use; rather, it coincided with unstructured and unverified engagement.

To test Research Question 2: *Does guided AI integration influence students' perceived learning, confidence, and satisfaction*, post-test composite comparisons demonstrated several significant differences favoring the Experimental Group. Students in the Experimental Group reported higher perceived global sourcing knowledge at post-test compared to the Control Group ($p < .05$). Students in the Experimental Group also reported significantly higher scores on the professional growth composite, including perceived communication development, application of skills to future contexts, and readiness for professional presentations ($p < .05$).

Both groups demonstrated increases in project confidence from pre- to post-test; however, the difference between groups at post-test was not statistically significant. This suggests that project completion enhanced confidence in both conditions, although AI integration was associated with stronger performance outcomes. Self-reported reflections indicated that students in the Experimental Group experienced greater efficiency ($M = 4.50$; $t = 3.10$; $p < .05$), lower stress ($M = 3.90$; $t = 2.70$; $p < .05$), and increased confidence in managing workload ($M = 4.20$; $t = 2.95$; $p < .05$). Control Group students reported higher stress levels and greater uncertainty in workload management. Overall, Research Question 2 findings indicate that guided AI integration positively influenced perceived learning, efficiency, and professional growth while maintaining comparable confidence levels.

For Research Question 3: *How does instructional AI policy shape students' academic integrity perceptions and ethical awareness*, both groups demonstrated high levels of ethical awareness at post-test, with strong agreement regarding institutional integrity policies and the unethical nature of unacknowledged AI use. However, policy framing shaped perceptions in meaningful ways. Control Group students were more likely to classify AI use categorically as "cheating," whereas Experimental Group students demonstrated a more differentiated understanding, distinguishing between ethical guided use and unethical undisclosed use. Indicators consistent with undisclosed AI assistance were observed primarily within the Control Group. This divergence between stated policy and observed engagement suggests that prohibition alone may not fully regulate AI use. These findings suggest that structured integration may promote ethical literacy more effectively than blanket bans.

Regarding Research Question 4: *What differences emerge in student-reported experiences between structured and prohibited AI conditions*, students in the Experimental Group reported greater enthusiasm toward AI as a learning tool ($p < .01$). Approximately 80% reported cross-checking AI outputs, and qualitative responses indicated primary use of AI for brainstorming, outlining, summarizing, and initial drafting, alongside active reflection on verification and human oversight. Thematic analysis revealed five dominant themes: AI as organizational scaffold, AI as idea generator, AI requiring verification, efficiency gains, and enhanced critical evaluation of sources. In contrast, the Control Group lacked structured engagement with AI tools yet exhibited evidence of informal AI consultation without consistent verification protocols. Presentation quality and engagement levels were generally lower in this group, with students appearing more stressed

and less organized during delivery. These differences suggest that policy context shapes not only outcomes but also the way students engage with AI tools.

## DISCUSSION AND RECOMMENDATIONS

The findings of this quasi-experimental study provide empirical insight into a central theoretical question: under what instructional conditions does generative AI function as a cognitive scaffold rather than a cognitive shortcut? By directly comparing structured AI integration with a prohibition-based condition within the same project-based course, this study contributes evidence relevant to cognitive load theory, scaffolding theory, and self-regulated learning frameworks.

Consistent with cognitive load theory (Sweller, 1988), the results indicate that guided AI use was associated with stronger written analytical performance and higher overall project scores in the Experimental Group. These differences were statistically significant, with large effect sizes observed for both written performance ($t(44) = 12.16$, $p < .001$, $d = 3.08$) and total project performance. Importantly, gains were most pronounced in dimensions requiring organization, synthesis, and conceptual alignment rather than simple content reproduction. This pattern suggests that when AI is positioned as a structured support tool, it may free working memory resources for deeper processing rather than replace higher-order reasoning. In contrast, the Control Group did not demonstrate comparable performance gains. Indicators consistent with AI-assisted writing were observed in some Control Group submissions; however, these observations were not formally quantified and are interpreted cautiously. Performance outcomes in the Control Group were lower on average, particularly in written analytical components. From a cognitive perspective, this pattern may reflect reduced germane processing: AI-generated content used without structured verification may create an "illusion of competence," as described in prior literature, without supporting durable conceptual integration. Thus, the findings align with the proposition that the cognitive consequences of AI depend not on access alone, but on instructional framing.

The results also provide applied support for scaffolding theory (Wood et al., 1976). In the Experimental Group, generative tools were embedded within explicit expectations, including documentation requirements, verification protocols, and instructor guidance emphasizing human oversight. Under these conditions, AI appears to have functioned as a temporary support mechanism, enabling students to perform complex tasks, such as strategic vendor analysis and synthesis of sourcing frameworks, while maintaining responsibility for evaluation and refinement. Students reported cross-checking outputs, revising drafts, and supplementing AI-generated material with independent research. This pattern reflects scaffolding rather than substitution: support was provided, but cognitive ownership remained with the learner. By contrast, in the absence of structured scaffolding, AI engagement in the Control Group appeared less regulated. The lack of formal guidance may have limited metacognitive monitoring and verification processes. In this sense, the findings are consistent with Kirschner et al.'s (2006) argument that minimally guided approaches may be less effective in complex learning environments.

Self-regulated learning theory (Zimmerman, 2002) further contextualizes these results. Students in the Experimental Group demonstrated evidence of planning, monitoring, and evaluating AI-generated outputs. The requirement to log AI use and verify content likely supported metacognitive engagement. In contrast, prohibition without structured engagement did not appear to foster comparable regulatory behaviors. Taken together, the findings suggest that generative AI is not inherently beneficial or harmful to learning. Rather, its cognitive function appears to be shaped by instructional design.

From a policy perspective, these results indicate that the effectiveness of AI governance should not be evaluated solely in terms of compliance. While prohibition may signal institutional commitment to academic integrity, it may not fully prevent AI engagement in environments where such tools are widely accessible. Within the context of this course, structured integration was associated with stronger performance outcomes and more differentiated ethical reasoning. These findings suggest that transparent integration frameworks, anchored in verification and instructional scaffolding, can coexist with academic rigor.

Future research should extend these findings using randomized designs and longitudinal measures to examine whether scaffolded AI use produces durable improvements in analytical skill development. Additionally, fine-grained analyses of student-AI interaction patterns may clarify how instructional guidance shapes cognitive processing over time.

In conclusion, the productive question for higher education is not whether students should use generative AI, but how instructional design can ensure that AI functions as scaffolding rather than substitution. By situating AI integration within established cognitive and instructional theory, this study provides an empirically grounded framework for moving beyond reactive policy debates toward principled pedagogical design.

**Limitations**

Several limitations should be acknowledged. First, the study was conducted within a single course and department, limiting generalizability to other disciplines or institutional contexts. Second, the quasi-experimental design relied on intact sections without random assignment; although baseline equivalence was established at pre-test, unmeasured group differences cannot be entirely ruled out. Third, project scores were assigned at the team level and applied to individual team members for statistical analysis. While this approach reflects course grading procedures, it may reduce independence of observations. In addition, grading was conducted by the course instructor, which may introduce potential rater effects despite use of a standardized rubric. Fourth, identification of AI-assisted writing in the Control Group was based on stylistic and structural indicators rather than systematic detection tools or independent coding procedures. These observations are therefore interpretive rather than quantifiable measures of AI use. Fifth, although survey measures demonstrated acceptable internal consistency, self-reported learning gains may not perfectly reflect objective knowledge acquisition. Finally, the modest sample size reduces statistical power for detecting smaller effects and may limit precision of effect size estimates.

Future research should consider larger multi-institution samples, independent rates for written evaluation, matched longitudinal designs, and systematic measurement of AI engagement patterns to strengthen causal inference.

## REFERENCES

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative AI: Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open, 11*, Article 101299. https://doi.org/10.1016/j.ssaho.2025.101299

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*(2), 369–385. https://doi.org/10.1007/s10734-022-00937-2

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta-systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating? Ensuring academic integrity in the era of ChatGPT. *Assessment & Evaluation in Higher Education, 48*(8), 1281–1296. https://doi.org/10.1080/02602938.2023.2190148

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, S., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. M., du Boulay, B., & Koedinger, K. (2022). Ethics in artificial intelligence in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 527–560. https://doi.org/10.1007/s40593-022-00279-w

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

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Korseberg, L., & Elken, M. (2025). Waiting for the revolution: How higher education institutions initially responded to ChatGPT. *Higher Education, 89*(4), 953–968. https://doi.org/10.1007/s10734-024-01256-4

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Naznin, K., Al Mahmud, A., Nguyen, M. T., & Chua, C. (2025). ChatGPT integration in higher education for personalized learning, academic writing, and coding tasks: A systematic review. *Computers, 14*(2), 53. https://doi.org/10.3390/computers14020053

Perkins, M. (2023). Academic integrity considerations of AI-generated text. *Journal of University Teaching & Learning Practice, 20*(2), Article 7. https://doi.org/10.53761/1.20.2.07

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–354. https://doi.org/10.37074/jalt.2023.6.1.9

Selwyn, N., & Abidoye, Y. (2018). The possibilities and limits of "techno-solutionism" in education: A critical analysis of artificial intelligence and datafication. *Learning, Media and Technology, 43*(3), 255–269. https://doi.org/10.1080/17439884.2018.1504797

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

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2