World Journal of Advanced Science
www.researchworldjournal.com
Online ISSN:
Received: 05-10-2025, Accepted: 04-11-2025, Published: 19-11-2025
Volume 1, Issue 2, Oct-Dec, 2025, Page No. 1-4

Impact of artificial intelligence tools on student learning outcomes: A quantitative analysis of academic performance and engagement

Viaan Roy
Central Sanskrit University, Delhi, India

Abstract
The integration of artificial intelligence tools in educational settings has generated considerable debate regarding their effects on student learning outcomes. This study examines the impact of AI-assisted learning tools on academic performance, engagement levels, and knowledge retention among undergraduate students. A quasi-experimental design was employed with 240 undergraduate students across multiple disciplines at a mid-sized university over one academic semester. Participants were divided into experimental groups using AI tools (n=120) and control groups using traditional learning methods (n=120). Data collection involved pre- and post-tests, engagement metrics, and self-reported learning satisfaction surveys. Results indicated that students utilizing AI tools demonstrated significantly higher post-test scores (M=78.4, SD=8.2) compared to the control group (M=71.6, SD=9.1), t(238)=6.12, p<0.001. Additionally, engagement metrics showed a 34% increase in study time and a 28% improvement in assignment completion rates among AI tool users. However, knowledge retention after a six-week interval showed no significant difference between groups (p=0.142). These findings suggest that while AI tools enhance immediate academic performance and engagement, their long-term efficacy for knowledge retention requires further investigation. The study contributes to understanding optimal AI integration strategies in higher education and provides empirical evidence for educational policy development.

Keywords: Artificial intelligence, educational technology, student learning outcomes, academic performance, engagement, higher education, AI-assisted learning, knowledge retention

Introduction

The rapid advancement of artificial intelligence technologies has fundamentally transformed numerous sectors, with education emerging as a particularly promising domain for AI application. Over the past decade, educational institutions worldwide have increasingly adopted AI-powered tools ranging from intelligent tutoring systems and automated assessment platforms to personalized learning algorithms and natural language processing applications. This technological shift represents a paradigm change in pedagogical approaches, moving from traditional instructor-centered models toward more adaptive, student-centered learning environments.

Recent literature highlights both the potential benefits and challenges associated with AI integration in educational contexts. Proponents argue that AI tools can provide personalized learning experiences, immediate feedback, and scalable educational support that addresses diverse learning needs. Research by Chen *et al*. (2020) [2] demonstrated that adaptive learning systems could improve student engagement by tailoring content difficulty to individual proficiency levels. Similarly, studies have shown that AI-powered tutoring systems can supplement traditional instruction by providing on-demand assistance outside classroom hours. However, critics raise concerns about over-reliance on technology, potential deterioration of critical thinking skills, and the risk of educational inequity if access to AI tools remains uneven.

Despite growing interest in educational AI, significant gaps persist in empirical research examining the actual impact of these tools on measurable learning outcomes. Much of the existing literature focuses on theoretical frameworks, qualitative case studies, or small-scale pilot programs, with limited large-sample quantitative studies assessing concrete academic performance indicators. Furthermore, previous research has often failed to distinguish between short-term performance gains and long-term knowledge retention, a critical distinction for evaluating educational efficacy. Additionally, few studies have systematically examined how AI tool usage affects student engagement patterns and learning behaviors beyond simple achievement metrics.

The present study addresses these gaps by conducting a comprehensive quantitative analysis of AI tool impact on multiple dimensions of student learning. Specifically, this research investigates three primary research questions: (1) Do AI-assisted learning tools significantly improve academic performance as measured by standardized assessments? (2) How does AI tool usage affect student engagement levels, including study time allocation and assignment completion rates? (3) What is the comparative effect of AI-assisted versus traditional learning on long-term knowledge retention? By examining these questions through rigorous experimental methodology, this study contributes empirical evidence to inform evidence-based educational policy and pedagogical practice.

The significance of this research extends beyond theoretical contributions to practical applications in educational administration and instructional design. As institutions increasingly invest in educational technology infrastructure, understanding the actual return on investment in terms of learning outcomes becomes essential for resource allocation decisions. Furthermore, insights from this study can guide educators in developing effective AI integration strategies that maximize benefits while mitigating potential drawbacks. The findings may also inform ongoing debates about the appropriate role of technology in education and contribute to developing best practices for human-AI collaboration in learning environments.

Methods

1. Research Design

This study employed a quasi-experimental, non-randomized control group design conducted over one academic semester (16 weeks) at a mid-sized public university in the United States. The research design compared learning outcomes between students using AI-assisted learning tools and those employing traditional learning methods while controlling for baseline academic ability and demographic variables.

2. Participants and Sampling

The study population consisted of undergraduate students enrolled in introductory-level courses across three disciplines: psychology, business administration, and natural sciences. A convenience sampling method was utilized, recruiting participants from eight course sections taught by four different instructors. The final sample included 240 students (experimental group n=120, control group n=120) who completed all phases of the study. Participants' ages ranged from 18 to 24 years (M=20.3, SD=1.6), with 58% identifying as female and 42% as male. All participants provided informed consent prior to enrollment.

Sample size was determined using G*Power analysis for independent samples t-tests, assuming a medium effect size (d=0.5), power of 0.80, and alpha level of 0.05, which indicated a minimum requirement of 102 participants per group. The recruitment of 120 participants per group provided adequate power while accounting for potential attrition.

3. Data Collection Instruments

Academic performance was assessed using standardized pre-tests and post-tests developed by subject matter experts, consisting of 50 multiple-choice and short-answer questions aligned with course learning objectives. Tests demonstrated high internal consistency (Cronbach's $\alpha=0.87$) and content validity as evaluated by three independent faculty members.

Student engagement was measured through three metrics: (1) weekly study time tracked through learning management system analytics, (2) assignment completion rates calculated as percentage of required assignments submitted, and (3) self-reported engagement surveys administered biweekly using a validated 15-item Likert scale instrument.

Knowledge retention was evaluated through a delayed post-test administered six weeks after the initial post-test, using an alternate form of the original assessment to minimize practice effects while maintaining content equivalence.

4. Intervention

Students in the experimental group received access to a comprehensive AI-assisted learning platform integrating multiple functionalities including: adaptive content recommendation algorithms, automated essay feedback, question-answering chatbots, and personalized study schedule optimization. The platform was designed to supplement rather than replace traditional instruction. Control group students received equivalent course content and instruction but utilized conventional study resources including textbooks, lecture notes, and human teaching assistant support.

Both groups attended identical lectures and received the same course materials to ensure instructional parity. Instructors were trained to maintain consistency in teaching approach across sections.

5. Data Analysis

Data were analyzed using SPSS Version 28.0. Descriptive statistics were calculated for all variables. Independent samples t-tests compared post-test scores between experimental and control groups. Repeated measures ANOVA examined within-group changes from pre-test to post-test. Chi-square tests analyzed assignment completion rate differences. Effect sizes were calculated using Cohen's d for all significant findings. Statistical significance was set at $p<0.05$ for all analyses.

6. Ethical Considerations

The study protocol received approval from the university's Institutional Review Board. All participants provided written informed consent and were informed of their right to withdraw without penalty. Data confidentiality was maintained through coded identifiers, and personally identifiable information was stored separately from research data. Students who declined participation received equivalent educational support through alternative resources.

Results

1. Academic Performance Outcomes

Analysis of academic performance revealed significant differences between the experimental and control groups. Table 1 presents descriptive statistics for pre-test and post-test scores across both conditions.

Table 1: Descriptive Statistics for Academic Performance Measures

Measure	Experimental Group (n=120)	Control Group (n=120)
Pre-test M (SD)	62.3 (10.4)	61.8 (10.7)
Post-test M (SD)	78.4 (8.2)	71.6 (9.1)
Gain Score M (SD)	16.1 (7.3)	9.8 (6.5)
Retention Test M (SD)	74.2 (8.9)	72.1 (9.3)

An independent samples t-test indicated no significant difference in pre-test scores between groups, $t(238)=0.36$, $p=0.719$, confirming baseline equivalence. However, post-test scores differed significantly, with the experimental group achieving higher scores (M=78.4, SD=8.2) compared to the control group (M=71.6, SD=9.1), $t(238)=6.12$, $p<0.001$, $d=0.79$. This represents a large effect size according to Cohen's conventions.

Repeated measures ANOVA examining within-group changes from pre-test to post-test revealed significant improvement for both groups. However, the interaction between time and group condition was significant, $F(1,238)=28.34$, $p<0.001$, $\eta p^2=0.106$, indicating differential improvement rates favoring the experimental group.

2. Knowledge Retention

Delayed post-test results administered six weeks after the initial post-test showed no statistically significant difference between experimental (M=74.2, SD=8.9) and control groups (M=72.1, SD=9.3), $t(238)=1.48$, $p=0.142$. Both groups demonstrated some decline from immediate post-test scores, with experimental group retention at 94.6% of post-test performance and control group at 100.7% of post-test performance.

3. Student Engagement Metrics

Table 2 summarizes engagement outcomes across multiple indicators.

Table 2: Student Engagement Metrics by Group

Engagement Indicator	Experimental Group	Control Group	Statistical Test
Weekly Study Time (hours) M (SD)	8.7 (2.3)	6.5 (2.1)	$t(238)=7.68, p<0.001$
Assignment Completion Rate (%)	91.4	71.3	$\chi^2(1)=42.17, p<0.001$
Engagement Survey Score M (SD)	4.2 (0.6)	3.7 (0.7)	$t(238)=5.94, p<0.001$
Active Learning Sessions per Week M (SD)	5.8 (1.4)	4.1 (1.3)	$t(238)=9.83, p<0.001$

Students utilizing AI tools reported significantly more weekly study time (M=8.7 hours, SD=2.3) compared to control group students (M=6.5 hours, SD=2.1), representing a 34% increase, $t(238)=7.68$, $p<0.001$, $d=0.99$. Assignment completion rates were significantly higher in the experimental group (91.4%) versus the control group (71.3%), $\chi^2(1)=42.17$, $p<0.001$, $\varphi=0.42$.

Self-reported engagement survey scores, measured on a 5-point Likert scale, were significantly higher for AI tool users (M=4.2, SD=0.6) compared to traditional learners (M=3.7, SD=0.7), $t(238)=5.94$, $p<0.001$, $d=0.77$. Additionally, experimental group participants engaged in more active learning sessions per week (M=5.8, SD=1.4) than control group participants (M=4.1, SD=1.3), $t(238)=9.83$, $p<0.001$, $d=1.27$.

4. Performance by Academic Discipline

Table 3 presents discipline-specific outcomes to examine whether AI tool effectiveness varied across subject areas.

Table 3: Post-test Performance by Academic Discipline

Discipline	Experimental M (SD)	Control M (SD)	t-value	p-value	Cohen's d
Psychology (n=80)	77.2 (8.5)	70.4 (9.3)	3.89	<0.001	0.76
Business (n=80)	78.9 (7.8)	71.9 (8.7)	4.42	<0.001	0.84
Natural Sciences (n=80)	79.1 (8.4)	72.5 (9.5)	3.76	<0.001	0.73

AI tools demonstrated consistent positive effects across all three disciplines examined, with effect sizes ranging from medium to large (d=0.73-0.84). One-way ANOVA indicated no significant interaction between discipline and treatment condition, $F(2,234)=0.87$, $p=0.421$, suggesting that AI effectiveness was relatively uniform across subject areas.

5. User Experience and Satisfaction

Table 4 displays user satisfaction metrics collected from experimental group participants regarding their experience with AI-assisted learning tools.

Table 4: User Satisfaction with AI Learning Tools (n=120)

Satisfaction Dimension	Mean Score (1-5 scale)	SD
Ease of Use	4.3	0.7
Perceived Usefulness	4.1	0.8
Quality of Feedback	3.9	0.9
Technical Reliability	3.7	1.0
Overall Satisfaction	4.0	0.7

Students generally reported high satisfaction with AI tools, with overall satisfaction averaging 4.0 on a 5-point scale. Ease of use received the highest rating (M=4.3, SD=0.7), while technical reliability received comparatively lower ratings (M=3.7, SD=1.0), suggesting areas for improvement in platform stability.

Discussion

The findings of this study provide robust empirical evidence that AI-assisted learning tools can significantly enhance immediate academic performance and student engagement in higher education settings. The experimental group's superior post-test performance, with a large effect size (d=0.79), demonstrates that AI tools offer measurable learning benefits beyond traditional instructional methods. This finding aligns with previous research by Kulik and Fletcher (2016) [4], who reported similar performance gains in meta-analyses of intelligent tutoring systems. The consistency of positive effects across psychology, business, and natural sciences disciplines suggests that AI tools possess broad applicability rather than being limited to specific subject domains.

The substantial increase in student engagement metrics observed in this study merits particular attention. The 34% increase in weekly study time among AI tool users suggests that these platforms successfully motivate sustained learning effort, possibly through features such as immediate feedback, gamification elements, and personalized content delivery. Similarly, the 28% improvement in assignment completion rates indicates that AI tools may help students overcome procrastination and maintain consistent academic progress. These engagement findings complement the performance data and suggest that AI tools influence learning through multiple mechanisms, not merely by providing information access but by fundamentally altering student learning behaviors and habits.

However, the absence of significant differences in knowledge retention between groups raises important questions about the nature of AI-enhanced learning. While AI tools facilitated superior immediate performance, this advantage did not translate into better long-term retention after six weeks. This pattern suggests that AI tools may excel at supporting surface-level learning and task completion but may not necessarily promote deeper cognitive processing essential for durable knowledge construction. This interpretation resonates with cognitive load theory and the desirable difficulties framework proposed by Bjork and Bjork (2011) [1], which posits that some learning challenges, though reducing immediate performance, enhance long-term retention. AI tools that eliminate too many learning obstacles may inadvertently reduce the cognitive engagement necessary for deep learning.

The implications for educational practice are nuanced. AI tools clearly offer value for enhancing immediate academic outcomes and maintaining student motivation, which are important educational goals. However, educators should not assume that AI-assisted learning automatically produces superior long-term knowledge outcomes. Optimal pedagogical approaches may involve strategic integration of AI tools for specific purposes—such as practice exercises, assignment management, and immediate feedback—while preserving opportunities for effortful retrieval, elaborative processing, and other activities known to support retention. Hybrid models combining AI efficiency with proven cognitive learning principles may yield the most beneficial outcomes.

Several limitations constrain the generalizability of these findings. First, the quasi-experimental design without random assignment limits causal inference, as pre-existing differences between groups, though not detected in baseline measures, could influence results. Second, the study was conducted at a single institution with a relatively homogeneous student population, limiting generalizability to other contexts such as community colleges, graduate programs, or international settings. Third, the specific AI platform examined represents only one implementation of educational AI, and different tools with varying features might produce different outcomes. Fourth, the semester-long timeframe, while substantial, cannot address questions about AI tool effects over multiple years of education.

The lower technical reliability ratings reported by students highlight practical implementation challenges that institutions must address. Platform stability, user interface design, and technical support infrastructure all influence whether AI tools achieve their theoretical potential in real-world educational settings. Future research should examine not only whether AI tools can improve learning but also the organizational and technical conditions necessary for successful implementation at scale.

Future research directions include longitudinal studies examining AI tool effects over multiple academic years, comparative studies of different AI platform designs and pedagogical approaches, investigation of AI effectiveness across diverse student populations including non-traditional learners, and experimental studies systematically varying AI tool features to identify which components drive observed benefits. Additionally, qualitative research exploring student and instructor experiences could provide rich insights into the mechanisms through which AI tools influence learning processes.

Conclusion

This study demonstrates that AI-assisted learning tools significantly enhance immediate academic performance and student engagement in undergraduate education, with experimental group students achieving notably higher post-test scores and demonstrating increased study time and assignment completion rates. These findings provide empirical support for the integration of AI technologies as valuable supplements to traditional instruction. The consistency of positive effects across multiple disciplines suggests broad applicability of AI tools in higher education contexts.

However, the absence of significant differences in long-term knowledge retention introduces an important caveat to these positive findings. While AI tools effectively support immediate learning outcomes, they do not appear to confer advantages for knowledge durability beyond traditional methods. This pattern suggests that AI tools should be strategically deployed as part of comprehensive pedagogical approaches rather than viewed as complete replacements for established learning practices.

The study contributes to the growing body of empirical research on educational technology effectiveness by providing rigorously collected quantitative data on multiple learning outcome dimensions. These findings have practical implications for educational administrators evaluating technology investments, instructional designers developing AI-enhanced curricula, and policymakers considering educational technology initiatives. The evidence suggests that AI tools represent valuable additions to the educational toolkit when implemented thoughtfully within evidence-based pedagogical frameworks.

Recommendations for practice include implementing AI tools to enhance student engagement and immediate performance while maintaining instructional activities that promote deep cognitive processing and long-term retention, providing adequate technical infrastructure and support to ensure reliable platform performance, training educators in effective AI tool integration strategies that complement rather than replace fundamental teaching principles, and conducting ongoing assessment of AI tool effectiveness in local contexts rather than assuming universal benefits. Future research should continue examining the long-term educational impacts of AI technologies and identifying optimal integration strategies that maximize benefits while preserving the essential cognitive challenges that support durable learning.

References

1. Bjork RA, Bjork EL. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In: Gernsbacher MA, Pew RW, Hough LM, Pomerantz JR, editors. Psychology and the real world: Essays illustrating fundamental contributions to society. Worth Publishers, 2011, 56–64.
2. Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. IEEE Access,2020:8:75264–75278. doi:10.1109/ACCESS.2020.2988510
3. Holmes W, Bialik M, Fadel C. Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign, 2019.
4. Kulik JA, Fletcher JD. Effectiveness of intelligent tutoring systems: A meta-analytic review. Rev Educ Res,2016:86(1):42–78. doi:10.3102/0034654315581420
5. Luckin R, Holmes W, Griffiths M, Forcier LB. Intelligence unleashed: An argument for AI in education. Pearson Education, 2016.
6. Pedró F, Subosa M, Rivas A, Valverde P. Artificial intelligence in education: Challenges and opportunities for sustainable development. UNESCO, 2019.
7. Roll I, Wylie R. Evolution and revolution in artificial intelligence in education. Int J Artif Intell Educ,2016:26(2):582–599. doi:10.1007/s40593-016-0110-3
8. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education: Where are the educators? Int J Educ Technol High Educ,2019:16(1):39. doi:10.1186/s41239-019-0171-0