DOI: 10.71317/jgst.2.8(s).2026.469
Journal of Global Social Transformation
Journal homepage: https://rjsaonline.org/index.php/JGST

# Generative AI, Bloom's Higher-Order Thinking Skills, and Academic Achievement: The Mediating Role of Student Engagement in Higher Education

**Sidra Saeed[1], Hina Shaukat[2], Muhammad Minhass Baloch[3], Laila Naz[4], Muhammad Aarab[5]**
[1] Faculty at Sound Foundation, Quetta
Email: Sidra.saeed52@yahoo.com
[2] Lecturer, Department of Educational Sciences, National University of Modern Languages
Email: hshoukat@numl.edu.pk
[3] Secondary School Teacher (SST), Department of Education, Balochistan
Email: mminhass1@gmail.com
[4] FG Fazaia Public School (2/S), E-9, Islamabad
Email: 1984hijabzahra@gmail.com
[5] Senior Scientific Officer, Department of Quality Control Laboratory, NIH, Islamabad
Email: M.aarab@nih.org.ok

**ARTICLE INFO**

**Submission:**
July 08, 2026
**Revised:**
July 24, 2026
**Accepted:**
August 09, 2026
**Available Online:**
August 23, 2026

**Keywords:**
*Generative AI, Bloom's Taxonomy, Higher-Order Thinking Skills, Student Engagement, Academic Achievement, Structural Equation Modeling, Higher Education.*

**Corresponding author:**
**Sidra Saeed**
**Email:**
Sidra.saeed52@yahoo.com

**ABSTRACT**

The rapid integration of generative artificial intelligence (GenAI) in higher education has recontextualized cognitive development, particularly regarding Bloom's higher-order thinking skills (HOTS). While GenAI's potential to enhance learning is widely recognized, the mechanisms through which GenAI-assisted HOTS influence academic achievement remain underexplored. This study addresses this gap by examining the mediating role of multidimensional student engagement (behavioral, emotional, and cognitive) in the relationship between GenAI-supported HOTS and academic achievement. Employing a quantitative, cross-sectional design, data were collected from 504 higher education students in Islamabad. Structural equation modeling (SEM) and bootstrap mediation analyses were utilized to test the hypothesized framework. Results indicated that GenAI-assisted HOTS significantly predicted both student engagement ($\beta = .62$, $p < .001$) and academic achievement ($\beta = .28$, $p < .001$). Crucially, student engagement partially mediated the relationship between GenAI-HOTS and academic achievement, accounting for 50.9% of the total effect ($\beta\_indirect = .29$, $p < .001$), with cognitive engagement emerging as the strongest mediating pathway. Multi-group analysis confirmed the structural invariance of this model across academic levels and disciplines. These findings demonstrate that GenAI's impact on academic achievement is significantly channeled through enhanced student engagement rather than occurring solely through direct cognitive offloading. The study provides empirical evidence for shifting pedagogical strategies from passive AI consumption to structured, engagement-driven AI integration that explicitly targets higher-order cognitive processes.

## Introduction

The rapid proliferation of generative artificial intelligence (GenAI) tools, such as large language models, AI-assisted writing platforms, and intelligent tutoring systems, has fundamentally disrupted the pedagogical landscape of higher education (Holmes & Tuomi, 2022). Since the public release of advanced generative models in late 2022, institutions worldwide have grappled with questions surrounding academic integrity, cognitive development, and the reconfiguration of learning outcomes (Cotton et al., 2024; Lim et al., 2023). While early discourse centered predominantly on the risks of plagiarism and cognitive offloading, a more nuanced scholarly conversation has emerged, examining how GenAI can be leveraged to cultivate higher-order cognitive processes and enhance academic performance (Yan et al., 2024). Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) provides a well-established hierarchical framework for classifying cognitive learning objectives, with the upper tiers, Analyzing, Evaluating, and Creating, constituting higher-order thinking skills (HOTS) that are essential for deep learning and intellectual autonomy (Krathwohl, 2002). In the context of GenAI integration, these higher-order skills assume heightened significance: students must critically evaluate AI-generated outputs, analyze the reasoning embedded within algorithmic responses, and synthesize novel contributions that transcend what the technology can produce independently (Mollick & Mollick, 2023; Sullivan et al., 2024).

Student engagement, broadly conceptualized as a multidimensional construct encompassing behavioral, emotional, and cognitive dimensions (Fredricks et al., 2004; Trowler, 2010), has long been recognized as a critical determinant of academic success in higher education (Kahu, 2013). Recent scholarship suggests that GenAI tools may influence engagement by altering the nature of student–content interaction, reducing cognitive load associated with lower-order tasks, and enabling more personalized learning pathways (Kasneci et al., 2023; Su & Yang, 2023). However, the precise mechanisms through which GenAI use influences engagement, and subsequently academic achievement, remain insufficiently theorized and empirically underexplored. Despite the accelerating adoption of GenAI in tertiary education, empirical research that simultaneously examines the interplay among GenAI usage, higher-order thinking skills, student engagement, and academic achievement within a unified analytical framework is conspicuously limited (Zawacki-Richter et al., 2024). Most existing studies adopt a fragmented approach, investigating isolated pairwise relationships rather than the complex mediating pathways that likely characterize these constructs (Bond et al., 2024). This gap is particularly pronounced in diverse higher education contexts where institutional policies, digital infrastructure, and student preparedness vary considerably.

The present study addresses this lacuna by proposing and empirically testing a structural model in which student engagement mediates the relationship between GenAI-assisted development of Bloom's higher-order thinking skills and academic achievement. By integrating cognitive taxonomy theory, engagement theory, and technology-enhanced learning frameworks, this research contributes to a more holistic understanding of how generative AI can be pedagogically harnessed to foster meaningful learning outcomes in higher education.

## Literature Review

### Generative AI in Higher Education: Opportunities and Pedagogical Implications

Generative AI refers to machine learning systems capable of producing novel text, images, code, and other content in response to user prompts (Cao et al., 2024). In higher education, tools such as ChatGPT, Claude, Gemini, and domain-specific AI assistants have been integrated into teaching, learning, and assessment practices at an unprecedented pace (Crawford et al., 2024). UNESCO (2023) reported that over 80% of surveyed universities acknowledged some form of GenAI interaction among students, prompting urgent calls for evidence-based pedagogical frameworks.

Scholars have identified several pedagogical affordances of GenAI, including scaffolding complex tasks, providing immediate formative feedback, enabling iterative drafting processes, and supporting differentiated instruction (Hwang & Chien, 2022; Luckin et al., 2023). Yan et al. (2024) demonstrated that structured GenAI integration in writing courses enhanced students' metacognitive awareness and prompted deeper engagement with content. Conversely, critical perspectives caution against over-reliance, arguing that unstructured AI use may erode foundational skills and promote superficial learning (Cotton et al., 2024; Dwivedi et al., 2023).

The pedagogical effectiveness of GenAI appears contingent upon instructional design. Mollick and Mollick (2023) found that when AI tools were embedded within structured learning activities requiring critical evaluation and creative extension, students demonstrated superior learning outcomes compared to unstructured use. This finding underscores the importance of aligning GenAI integration with explicit cognitive objectives, particularly those targeting higher-order thinking.

### Bloom's Higher-Order Thinking Skills: Theoretical Foundations and Contemporary Relevance

Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) reorganized the original six-level cognitive hierarchy into a two-dimensional framework comprising the knowledge dimension and the cognitive process dimension. The upper three levels—*Analyze* (breaking material into constituent parts and detecting relationships), *Evaluate* (making judgments based on criteria and standards), and *Create* (putting elements together to form a novel, coherent whole)—constitute higher-order thinking skills (HOTS) that demand complex cognitive processing beyond mere recall or comprehension (Krathwohl, 2002).

In higher education, HOTS are widely regarded as essential learning outcomes that prepare graduates for complex professional environments and lifelong learning (Biggs & Tang, 2011). Facione (2020) emphasized that analytical and evaluative thinking constitute the core of critical thinking dispositions necessary in knowledge economies. Similarly, creative thinking—the apex of Bloom's hierarchy—has been linked to innovation capacity and problem-solving efficacy (Runco & Jaeger, 2012).

The advent of GenAI has recontextualized HOTS within educational practice. Sullivan et al. (2024) argued that AI tools can handle lower-order cognitive tasks (remembering, understanding), thereby freeing cognitive resources for students to engage in analysis, evaluation, and creation. Kasneci et al. (2023) proposed that effective AI-augmented pedagogy should deliberately shift instructional focus toward HOTS, using AI as a catalyst rather than a substitute for deep thinking. Empirical evidence supports this view: Lim et al. (2023) found that students who used AI tools within an evaluative framework (e.g., critiquing AI outputs against disciplinary standards) demonstrated significantly higher analytical reasoning scores than those who used AI passively.

However, the relationship between GenAI use and HOTS is not inherently positive. Dwivedi et al. (2023) warned that without deliberate pedagogical structuring, GenAI may inadvertently encourage cognitive dependency, reducing opportunities for students to independently engage in analysis, evaluation, and creation. This tension highlights the need for mediating variables—such as student engagement—that may explain when and how GenAI use translates into higher-order cognitive development.

**Student Engagement: A Multidimensional Mediator**

Student engagement is a multifaceted construct that has been operationalized across behavioral, emotional, and cognitive dimensions (Fredricks et al., 2004). Behavioral engagement refers to participation and effort in academic activities; emotional engagement encompasses interest, belonging, and affective responses to learning; and cognitive engagement involves investment in deep learning strategies, self-regulation, and intellectual persistence (Trowler, 2010).

In technology-enhanced learning environments, engagement has been identified as a critical mediator between instructional interventions and learning outcomes (Bond et al., 2024; Martin et al., 2022). Kahu's (2013) psycho-sociological framework posits that engagement emerges from the interaction of institutional, curricular, and psychosocial factors, positioning it as a dynamic process rather than a static trait. Applied to GenAI contexts, this framework suggests that engagement mediates the extent to which AI tools translate into meaningful cognitive gains.

Recent empirical studies have begun exploring AI-specific engagement. Su and Yang (2023) found that students who interacted with AI tools in a self-regulated, inquiry-driven manner reported higher cognitive engagement and deeper processing of course material. Similarly, Hwang et al. (2024) demonstrated that AI-supported collaborative learning environments enhanced behavioral and emotional engagement, which in turn predicted improved problem-solving performance. Conversely, passive or purely instrumental AI use (e.g., generating answers without reflection) was associated with disengagement and surface-level learning (Cotton et al., 2024).

The mediating role of engagement between technology use and academic outcomes is well-established in broader educational technology literature. Martin et al. (2022) conducted a meta-analysis confirming that student engagement significantly mediates the relationship between digital tool integration and academic performance across 47 studies. Extending this logic to GenAI and HOTS, it is plausible that engagement serves as the mechanism through which AI-assisted higher-order thinking activities translate into tangible achievement gains.

**Academic Achievement in Higher Education**

Academic achievement in higher education is typically operationalized through grade point averages (GPA), course-specific performance metrics, standardized assessments, and competency-based evaluations (Richardson, 2005; Schneider & Preckel, 2017). While traditional metrics capture knowledge acquisition, contemporary scholarship advocates for multidimensional assessments that incorporate higher-order cognitive performance, creative outputs, and applied problem-solving (Biggs & Tang, 2011).

The relationship between HOTS and academic achievement is well-documented. Students who engage in analytical, evaluative, and creative thinking processes consistently outperform peers who rely predominantly on memorization and surface learning strategies (Entwistle & McCune, 2004; Richardson, 2005). Furthermore, engagement has been identified as a robust predictor of academic achievement, with cognitive engagement showing the strongest association with deep learning outcomes (Kahu, 2013; Trowler, 2010).

In the GenAI context, academic achievement is being redefined. Zawacki-Richter et al. (2024) noted that institutions are increasingly incorporating AI literacy, critical evaluation of AI outputs, and AI-augmented creative projects into assessment frameworks. This evolution suggests that academic achievement in AI-integrated environments may reflect not only content mastery but also the capacity to leverage technology for higher-order cognitive tasks.

**The Mediating Role of Student Engagement: An Integrative Framework**

Integrating the foregoing literature, a theoretical pathway emerges: GenAI tools, when pedagogically structured to target Bloom's higher-order thinking skills, may enhance student engagement (particularly cognitive and behavioral engagement), which in turn facilitates improved academic achievement. This mediating pathway aligns with the Technology-Enhanced Learning Engagement Model (TEL-EM; Martin et al., 2022) and the cognitive load framework (Sweller et al., 2019), which suggests that by offloading lower-order cognitive demands to AI, students can redirect attentional resources toward deeper engagement with higher-order tasks.

However, this mediation is not automatic. It is contingent upon instructional design, student self-regulation, institutional support, and the nature of AI interaction (Bond et al., 2024; Kasneci et al., 2023). Without structured engagement opportunities, GenAI use may bypass higher-order thinking entirely, resulting in superficial learning and diminished achievement (Dwivedi et al., 2023).

Despite the theoretical plausibility of this mediating model, empirical validation remains limited. Most studies examine GenAI and engagement, or engagement and achievement, in isolation (Zawacki-Richter et al., 2024). Few have employed structural equation modeling or mediation analysis to test the indirect pathway from GenAI-assisted HOTS development through engagement to academic achievement. The present study seeks to address this gap.

**Research Objectives**

The present study is guided by the following objectives:

1. To examine the extent to which the use of generative AI tools in higher education is associated with the development of Bloom's higher-order thinking skills (Analyzing, Evaluating, and Creating) among university students.
2. To investigate the relationship between generative AI-supported higher-order thinking skills and student engagement (behavioral, emotional, and cognitive dimensions) in higher education settings.
3. To assess the direct and indirect effects of generative AI-assisted higher-order thinking skills on academic achievement, with student engagement as a mediating variable.
4. To determine the mediating role of student engagement in the relationship between GenAI-supported higher-order thinking skills and academic achievement using structural equation modeling.
5. To explore whether the mediating effect of student engagement varies across demographic and academic subgroups (e.g., discipline, year of study, prior AI experience).
6. To provide evidence-based pedagogical recommendations for integrating generative AI in higher education curricula to optimize higher-order thinking, engagement, and academic outcomes.

## Methodology

**Research Design**

The present study employed a quantitative, cross-sectional research design grounded in a positivist paradigm. A structured survey methodology was adopted to collect primary data from higher education students in Islamabad. The study utilized a structural equation modeling (SEM) approach to test the hypothesized mediation model in which student engagement mediates the relationship between GenAI-supported higher-order thinking skills and academic achievement. The cross-sectional design was deemed appropriate for examining concurrent relationships among the study variables and testing the proposed theoretical model (Creswell & Creswell, 2023).

**Population and Sampling**

The target population comprised undergraduate and postgraduate students enrolled in higher education institutions. A multi-stage stratified random sampling technique was employed to ensure representation across disciplines (Sciences, Social Sciences, Humanities, and Professional Studies) and academic levels (undergraduate and postgraduate).

The minimum sample size was determined using the Monte Carlo simulation method for SEM, which recommends a minimum of 200-300 observations for models with four latent constructs and approximately 20 indicators (Wolf et al., 2013). Additionally, the rule of 10 observations per estimated parameter (Kline, 2023) was applied.

**Inclusion Criteria**

Participants were included if they: (a) were currently enrolled in a higher education institution; (b) had used at least one generative AI tool (e.g., ChatGPT, Gemini, Claude, Copilot, or equivalent) for academic purposes within the preceding academic semester; and (c) provided informed consent to participate.

**Instruments**

A structured questionnaire comprising five sections was developed, drawing on validated scales from existing literature. All items were measured on a 5-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*) unless otherwise specified.

**Generative AI Usage and Higher-Order Thinking Skills (GenAI-HOTS)**

A 12-item scale was adapted from the Revised Bloom's Taxonomy framework (Anderson & Krathwohl, 2001) and the AI-Integrated Critical Thinking Scale (Lim et al., 2023). The scale comprised three sub-dimensions:

- *Analyzing* (4 items; e.g., "I use GenAI outputs to compare and contrast different perspectives on a topic.")
- *Evaluating* (4 items; e.g., "I critically assess the accuracy and reliability of GenAI-generated information before using it.")
- *Creating* (4 items; e.g., "I use GenAI as a starting point to develop original ideas and solutions.")

**Student Engagement**

A 12-item scale was adapted from the Student Engagement Instrument (SEI; Fredricks et al., 2004) and the Online Student Engagement Scale (Martin et al., 2022), modified for AI-enhanced learning contexts. The scale comprised three sub-dimensions:

- *Behavioral Engagement* (4 items; e.g., "I actively participate in learning activities that involve GenAI tools.")
- *Emotional Engagement* (4 items; e.g., "I feel motivated and interested when using GenAI to explore course material.")
- *Cognitive Engagement* (4 items; e.g., "I invest effort in deeply understanding concepts when GenAI assists my learning.")

**Academic Achievement**

Academic achievement was measured using a 6-item self-reported scale adapted from Richardson (2005) and Schneider and Preckel (2017). Items assessed perceived academic performance, grade attainment relative to peers, and achievement of course learning outcomes (e.g., "My overall academic performance this semester reflects my understanding of higher-order concepts."). Additionally, participants reported their cumulative Grade Point Average (GPA) on a standardized scale

**Demographic and GenAI Usage Profile**

Participants provided information on age, gender, academic level, discipline, and GenAI usage patterns (frequency, duration, purpose, and specific tools used).

**Validity and Reliability**

Content validity was established through expert review by five subject matter experts in educational technology, cognitive psychology, and higher education pedagogy. A Content Validity Index (CVI) of 0.92 was achieved, exceeding the recommended threshold of 0.80 (Polit & Beck, 2006). Construct validity was assessed through Confirmatory Factor Analysis (CFA). Internal consistency was evaluated using Cronbach's alpha ($\alpha \geq 0.70$) and Composite Reliability (CR $\geq 0.70$). Discriminant validity was assessed using the Fornell-Larcker criterion and the Heterotrait-Monotrait (HTMT) ratio (Henseler et al., 2015).

**Data Collection Procedure**

Data were collected over a period of eight weeks using an online survey platform (Qualtrics). The questionnaire was distributed through institutional email lists, learning management systems, and departmental coordinators. Participants were provided with an information sheet detailing the study's purpose, voluntary nature, and data confidentiality protocols. Informed consent was obtained electronically prior to survey administration. The average completion time was approximately 18–22 minutes.

## Results

**Response Rate and Data Screening**

A total of 547 responses were received. After removing 31 incomplete responses and 12 multivariate outliers (Mahalanobis distance, $p < .001$), a final sample of 504 valid responses was retained for analysis. The effective response rate was 92.1%. Skewness values ranged from −0.82 to 0.64, and kurtosis values ranged from −0.91 to 1.12, indicating acceptable normality (Kline, 2023). Variance Inflation Factor (VIF) values ranged from 1.12 to 2.87, confirming the absence of multicollinearity.

**Demographic Profile**

**Table 1** presents the demographic characteristics of the respondents.

***Table 1: Demographic Profile of Respondents (N = 504)***

| Characteristic | Category | *N* | % |
|---|---|---|---|
| Gender | Male | 231 | 45.8 |
| | Female | 258 | 51.2 |
| | Prefer not to say | 15 | 3.0 |
| Age (years) | 18–21 | 198 | 39.3 |
| | 22–25 | 176 | 34.9 |
| | 26–30 | 87 | 17.3 |
| | Above 30 | 43 | 8.5 |
| Academic Level | Undergraduate | 312 | 61.9 |
| | Postgraduate | 192 | 38.1 |
| Discipline | Sciences & Engineering | 168 | 33.3 |
| | Social Sciences | 142 | 28.2 |
| | Humanities & Arts | 97 | 19.2 |
| | Professional Studies (Business, Law, Medicine) | 97 | 19.2 |
| GenAI Usage Frequency | Daily | 89 | 17.7 |
| | Several times a week | 196 | 38.9 |
| | Once a week | 124 | 24.6 |
| | A few times a month | 95 | 18.8 |
| Primary GenAI Tools Used | ChatGPT / GPT-4 | 387 | 76.8 |
| | Google Gemini | 214 | 42.5 |
| | Claude | 128 | 25.4 |
| | Microsoft Copilot | 112 | 22.2 |
| | Other / Multiple tools | 63 | 12.5 |

*Note.* Percentages for GenAI tools exceed 100% as respondents could select multiple tools.

The sample of 504 students is well-balanced in gender (slight female majority at 51.2%) and predominantly consists of young adults, with nearly 75% aged between 18 and 25. The majority are undergraduate students (61.9%), but there is strong postgraduate representation (38.1%). The sample is well-distributed across all major fields of study (Sciences & Engineering, Social Sciences, Humanities & Arts, and Professional Studies), ensuring the findings are not biased toward a single discipline. GenAI use is highly regular. Over 80% of students use these tools at least once a week, with "several times a week" being the most common (38.9%) and 17.7% using it daily. ChatGPT/GPT-4 is the overwhelmingly dominant tool (76.8%), followed by Google Gemini (42.5%).

**Descriptive Statistics and Correlation Analysis**

**Table 2** presents the means, standard deviations, and Pearson correlation coefficients among the study variables.

**Table 2: Descriptive Statistics and Pearson Correlations (N = 504)**

| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| 1. GenAI-Analyzing | 3.72 | 0.71 | 1 | | | | | | |
| 2. GenAI-Evaluating | 3.65 | 0.74 | .68** | 1 | | | | | |
| 3. GenAI-Creating | 3.58 | 0.79 | .64** | .67** | 1 | | | | |
| 4. Behavioral Engagement | 3.81 | 0.68 | .52** | .54** | .49** | 1 | | | |
| 5. Emotional Engagement | 3.69 | 0.73 | .47** | .50** | .46** | .71** | 1 | | |
| 6. Cognitive Engagement | 3.74 | 0.70 | .56** | .58** | .53** | .69** | .72** | 1 | |
| 7. Academic Achievement | 3.78 | 0.65 | .48** | .51** | .45** | .55** | .52** | .59** | 1 |

*Note.* ** $p < .01$ (two-tailed). GenAI = Generative Artificial Intelligence. All variables measured on a 5-point Likert scale.

All inter-construct correlations were statistically significant ($p < .01$) and ranged from .45 to .72, indicating moderate to strong positive relationships. No correlation exceeded .85, suggesting the absence of problematic multicollinearity (Kline, 2023). All variables have means ranging from 3.58 to 3.81 on a 5-point scale, indicating that students generally reported high levels of GenAI use for higher-order thinking, strong student engagement, and positive perceptions of their academic achievement. Every variable is positively and significantly correlated with every other variable ($p < .01$). This means that as students' use of GenAI for higher-order thinking increases, their behavioral, emotional, and cognitive engagement, as well as their academic achievement, also tend to increase. The table confirms that GenAI-assisted higher-order thinking, student engagement, and academic achievement are all strongly and positively linked, setting a solid foundation for the structural equation modeling (SEM) that follows.

**Measurement Model: Confirmatory Factor Analysis**

A Confirmatory Factor Analysis (CFA) was conducted to assess the psychometric properties of the measurement model. All factor loadings exceeded the recommended threshold of 0.70, ranging from 0.72 to 0.89.

**Table 3: Confirmatory Factor Analysis: Factor Loadings, Reliability, and Validity**

| Construct / Item | Factor Loading (λ) | Cronbach's α | CR | AVE |
|---|---|---|---|---|
| **GenAI-Analyzing** | | | | |
| GA1 | .76 | | | |
| GA2 | .78 | .84 | .85 | .58 |
| GA3 | .74 | | | |
| GA4 | .77 | | | |
| **GenAI-Evaluating** | | | | |
| GE1 | .80 | | | |
| GE2 | .79 | .86 | .87 | .62 |
| GE3 | .77 | | | |
| GE4 | .79 | | | |
| **GenAI-Creating** | | | | |
| GC1 | .75 | | | |
| GC2 | .78 | .85 | .86 | .60 |
| GC3 | .79 | | | |
| GC4 | .78 | | | |
| **Behavioral Engagement** | | | | |
| BE1 | .74 | | | |
| BE2 | .76 | .83 | .84 | .57 |
| BE3 | .77 | | | |

| | | | | |
|---|---|---|---|---|
| BE4 | .75 | | | |
| **Emotional Engagement** | | | | |
| EE1 | .78 | | | |
| EE2 | .76 | .85 | .86 | .60 |
| EE3 | .79 | | | |
| EE4 | .76 | | | |
| **Cognitive Engagement** | | | | |
| CE1 | .77 | | | |
| CE2 | .78 | .84 | .85 | .59 |
| CE3 | .76 | | | |
| CE4 | .78 | | | |
| **Academic Achievement** | | | | |
| AA1 | .81 | | | |
| AA2 | .79 | .87 | .88 | .63 |
| AA3 | .80 | | | |
| AA4 | .78 | | | |
| AA5 | .77 | | | |
| AA6 | .79 | | | |

*Note.* CR = Composite Reliability; AVE = Average Variance Extracted. All factor loadings significant at $p < .001$.

Table 3 confirms the robust psychometric properties of the measurement model via Confirmatory Factor Analysis (CFA). All factor loadings (ranging from 0.74 to 0.81) exceeded the 0.70 threshold, indicating that the survey items strongly and accurately represent their underlying constructs. The model demonstrated excellent internal consistency, with Cronbach's alpha and Composite Reliability (CR) values ranging from 0.83 to 0.88, well above the standard 0.70 benchmark. Furthermore, the Average Variance Extracted (AVE) values (0.57 to 0.63) surpassed the 0.50 threshold, confirming strong convergent validity. Overall, these results prove that the instruments used to measure GenAI-assisted higher-order thinking, student engagement, and academic achievement are highly reliable and valid for the study. All constructs demonstrated satisfactory internal consistency (α range: 0.83–0.87), composite reliability (CR range: 0.84–0.88), and convergent validity (AVE range: 0.57–0.63), exceeding recommended thresholds (Hair et al., 2022).

**Model Fit Indices**

**Table 04. Goodness-of-Fit Indices for Measurement and Structural Models**

| Fit Index | Recommended Threshold | Measurement Model (CFA) | Structural Model (SEM) |
|---|---|---|---|
| $\chi^2$ (*df*) | — | 612.48 (384) | 648.72 (392) |
| $\chi^2$/df | < 3.00 | 1.59 | 1.65 |
| CFI | > 0.95 | 0.971 | 0.968 |
| TLI | > 0.95 | 0.966 | 0.963 |
| RMSEA (90% CI) | < 0.06 | 0.034 (0.029–0.039) | 0.036 (0.031–0.041) |
| SRMR | < 0.08 | 0.038 | 0.042 |
| GFI | > 0.90 | 0.931 | 0.927 |
| AGFI | > 0.90 | 0.912 | 0.908 |

*Note. $\chi^2$ = Chi-square; df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index. Both models demonstrate excellent fit (Hu & Bentler, 1999).*

Table 4 demonstrates that both the measurement (CFA) and structural (SEM) models exhibit an excellent fit to the observed data. All key goodness-of-fit indices comfortably met or exceeded the recommended statistical thresholds: the relative chi-square ($\chi^2$/df) was well below 3.00, incremental fit indices (CFI and TLI) were above 0.95, and absolute fit indices (RMSEA < 0.06, SRMR < 0.08, GFI and AGFI > 0.90) were all well within acceptable limits. This confirms that the hypothesized relationships and the underlying measurement structure are statistically robust, providing a highly valid foundation for the subsequent path and mediation analyses.Structural Model: Path Analysis The structural model was estimated to test the hypothesized direct and indirect relationships. Table 5 presents the standardized path coefficients, standard errors, critical ratios, and significance levels.

**Structural Model**

***Table 5: Structural Model: Standardized Path Coefficients***

| Path (Hypothesis) | β | *SE* | *CR* |
|---|---|---|---|
| GenAI-HOTS → Student Engagement (H1) | .62 | .04 | 14.87 |
| GenAI-HOTS → Academic Achievement (H2) | .28 | .05 | 5.91 |
| Student Engagement → Academic Achievement (H3) | .47 | .05 | 9.63 |
| GenAI-Analyzing → Student Engagement | .24 | .04 | 5.72 |
| GenAI-Evaluating → Student Engagement | .22 | .04 | 5.34 |
| GenAI-Creating → Student Engagement | .19 | .04 | 4.68 |
| Behavioral Engagement → Academic Achievement | .18 | .04 | 4.21 |
| Emotional Engagement → Academic Achievement | .14 | .04 | 3.45 |

*Note.* β = standardized regression coefficient; *SE* = standard error; *CR* = critical ratio; CI = confidence interval. GenAI-HOTS = Composite higher-order thinking skills construct. The structural model explained 58.4% of variance in Student Engagement ($R^2$ = .584) and 52.7% of variance in Academic Achievement ($R^2$ = .527).

The structural model confirms that all hypothesized direct pathways are statistically significant and positive. GenAI-assisted Higher-Order Thinking Skills (HOTS) strongly predict Student Engagement (β = .62) and also have a significant direct impact on Academic Achievement (β = .28). Additionally, Student Engagement is a robust predictor of Academic Achievement (β = .47). At the sub-dimensional level, the "Analyzing" component of HOTS is the strongest driver of student engagement, while "Behavioral Engagement" is the strongest predictor of academic achievement among the dimensions shown. All critical ratios (CR) are well above the threshold for statistical significance, demonstrating the robustness of these relationships and validating the study's core hypotheses.

**Mediation Analysis**

Bootstrap mediation analysis (5,000 resamples; 95% bias-corrected confidence intervals) was conducted to test the indirect effect of GenAI-HOTS on Academic Achievement through Student Engagement.

***Table 6. Mediation Analysis: Direct, Indirect, and Total Effects***

| Effect | B | *SE* | 95% CI (Lower) | 95% CI (Upper) | *P* |
|---|---|---|---|---|---|
| Total Effect: GenAI-HOTS → Academic Achievement | .57 | .04 | .49 | .65 | < .001 |
| Direct Effect: GenAI-HOTS → Academic Achievement | .28 | .05 | .19 | .38 | < .001 |
| Indirect Effect: GenAI-HOTS → SE → Academic Achievement | .29 | .03 | .23 | .35 | < .001 |
| Proportion Mediated | 50.9% | — | — | — | — |
| Specific Indirect Effects: | | | | | |
| GenAI-HOTS → Behavioral Eng. → Achievement | .11 | .02 | [.07, .15] | < .001 | |
| GenAI-HOTS → Emotional Eng. → Achievement | .09 | .02 | [.05, .13] | < .001 | |
| GenAI-HOTS → Cognitive Eng. → Achievement | .12 | .02 | [.08, .16] | < .001 | |

*Note. SE = Student Engagement (composite mediator). Bootstrap samples = 5,000. CI = bias-corrected confidence interval. The indirect effect is significant as the 95% CI does not include zero. Student Engagement partially mediates the relationship between GenAI-HOTS and Academic Achievement, accounting for 50.9% of the total effect.*

The bootstrap mediation analysis confirms that Student Engagement acts as a significant partial mediator between GenAI-assisted Higher-Order Thinking Skills (HOTS) and Academic Achievement. While the overall (total) effect of HOTS on achievement is strong (β = .57), the results show that 50.9% of this impact is channeled indirectly through enhanced student engagement (indirect effect β = .29), while the remaining effect is direct (β = .28). Furthermore, the specific indirect effects reveal that Cognitive Engagement (β = .12) is the strongest mediating pathway, followed closely by Behavioral (β = .11) and Emotional Engagement (β = .09). Because the 95% confidence intervals for all indirect effects exclude zero, the mediation is statistically robust, proving that GenAI significantly boosts academic success by actively engaging students in the learning process.The results indicate partial mediation: Student Engagement significantly mediates the relationship between GenAI-supported HOTS and Academic Achievement ($\beta\_indirect = .29$, $p < .001$), while a significant direct effect also persists ($\beta\_direct = .28$, $p < .001$). Among the three engagement dimensions, Cognitive Engagement demonstrated the strongest indirect pathway, followed by Behavioral Engagement and Emotional Engagement.

**Multi-Group Analysis (MGA)**

Multi-group analysis was conducted to examine whether the mediation model was invariant across academic level and discipline.

***Table 7. Multi-Group Analysis: Model Invariance Across Academic Level and Discipline***

| **Model Constraint** | **$\Delta\chi^2$** | **$\Delta df$** | ***p*** | **CFI Difference** |
|---|---|---|---|---|
| Academic Level (UG vs. PG) | | | | |
| Unconstrained (Configural) | — | — | — | — |
| Metric Invariance | 8.42 | 6 | .209 | .002 |
| Scalar Invariance | 12.67 | 12 | .394 | .003 |
| Discipline (4 groups) | | | | |
| Unconstrained (Configural) | — | — | — | — |
| Metric Invariance | 14.23 | 18 | .718 | .001 |
| Scalar Invariance | 21.56 | 36 | .971 | .002 |

UG = Undergraduate; PG = Postgraduate. $\Delta\chi^2$ < critical value and $p > .05$ indicate that constraints do not significantly worsen model fit, supporting measurement invariance across groups (Byrne, 2016).

Table 7 confirms that the proposed mediation model is highly robust and consistent across diverse student populations. The multi-group analysis successfully established both metric and scalar invariance across academic levels (undergraduate vs. postgraduate) and across four different academic disciplines. Because all *p*-values for the chi-square differences are well above .05 and the CFI differences are negligible (≤ .003, far below the strict .01 threshold), the model's fit is not significantly worsened by these constraints. In practical terms, this means the positive pathway from GenAI-assisted higher-order thinking, through student engagement, to academic achievement operates in the same reliable and predictable way, regardless of a student's year of study or field of specialization.

**Path Coefficient Comparison**

***Table 8. Path Coefficient Comparison Across Academic Levels***

| **Path** | **Undergraduate (β)** | **Postgraduate (β)** | **Δβ** |
|---|---|---|---|
| GenAI-HOTS → Student Engagement | .64 | .59 | .05 |
| GenAI-HOTS → Academic Achievement | .31 | .24 | .07 |
| Student Engagement → Academic Achievement | .44 | .51 | .07 |
| Indirect Effect (HOTS → SE → AA) | .28 | .30 | .02 |

Table 8 demonstrates that the underlying mechanisms of the model operate almost identically for both undergraduate and postgraduate students. While there are minor, statistically non-significant variations, such as GenAI-assisted higher-order thinking having a slightly stronger direct impact on engagement for undergraduates (β = .64 vs. .59), and engagement having a slightly stronger impact on academic achievement for postgraduates (β = .51 vs. .44), the overall pathways remain highly consistent. Most importantly, the core indirect (mediated) effect is virtually identical across both groups (Δβ = .02). This confirms that the pedagogical benefit of using GenAI to foster higher-order thinking, which in turn drives engagement and academic success, is a universal phenomenon regardless of a student's academic level.

## Conclusion

This study demonstrates that the pedagogical integration of generative AI in higher education significantly enhances academic achievement, primarily by fostering Bloom's higher-order thinking skills (HOTS) and driving multidimensional student engagement. Crucially, student engagement acts as a powerful partial mediator, accounting for over half (50.9%) of the total effect of GenAI-assisted HOTS on academic success, with *cognitive engagement* emerging as the strongest pathway. Furthermore, multi-group analysis confirms that this positive dynamic is robust and consistent across diverse academic levels and disciplines. Ultimately, the findings suggest that GenAI is not merely a tool for cognitive offloading, but a potent catalyst for deep learning, provided its use is intentionally structured to promote active analysis, evaluation, and creation.

## Recommendations

1. **For Educators (Pedagogical Design):** Shift assessment and assignment designs away from tasks susceptible to passive AI generation. Instead, create "AI-in-the-loop" activities that require students to critically evaluate, refine, and synthesize AI-generated outputs, thereby explicitly targeting analytical, evaluative, and creative thinking.
2. **For Institutions (Policy & Training):** Move beyond restrictive, plagiarism-focused AI policies. Develop comprehensive, discipline-agnostic AI literacy programs that train both students and faculty on *how* to use GenAI to deepen cognitive engagement and self-regulated learning.
3. **For Curriculum Developers:** Embed structured reflection prompts into AI-assisted tasks (e.g., "Explain how you verified the AI's output" or "Describe how you expanded upon the AI's initial idea") to deliberately trigger and sustain cognitive and emotional engagement.
4. **For Future Research:** Conduct longitudinal or experimental studies to establish causal relationships over time. Additionally, future research should investigate how specific training in "prompt engineering" for critical inquiry impacts the cognitive engagement of diverse student subgroups.

## References


1. 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.
2. Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
3. Bond, M., Khosravi, H., De Laat, M., Jovanović, N., Whitelock-Wainwright, A., Guan, Q., ... & Gašević, D. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *British Journal of Educational Technology*, *55*(4), 1391–1424.
4. Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2024). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *Frontiers of Computer Science*, *18*(3), 183302.
5. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239.
6. Crawford, J., Cowirka, M., Smith, C., & Butler-Henderson, K. (2024). Generative AI and the future of higher education: A thematic review. *Journal of Applied Learning and Teaching*, *7*(1), 1–18.
7. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642.
8. Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy and approaches. *Educational Psychology Review*, *16*(1), 1–47.
9. Facione, P. A. (2020). *Critical thinking: What it is and why it counts*. Insight Assessment.
10. Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109.

11. Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570.

12. Hwang, G.-J., & Chien, S.-Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, *3*, 100082.

13. Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2024). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100001.

14. Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758–773.

15. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274.

16. Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218.

17. Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790.

18. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2023). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

19. Martin, F., Sun, T., & Westine, C. (2022). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education*, *159*, 104009.

20. Mollick, E., & Mollick, L. (2023). *Using AI to implement effective teaching strategies in classrooms: Five strategies, with prompts*. Wharton Research Paper. University of Pennsylvania.

21. Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, *30*(4), 387–415.

22. Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96.

23. Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600.

24. Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355–366.

25. Sullivan, M., Kelly, A., & McLaughlan, P. (2024). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, *7*(1), 1–10.

26. Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292.

27. Trowler, V. (2010). *Student engagement literature review*. The Higher Education Academy.

28. UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

29. Yan, L., Sha, L., Zhao, L., Li, Y., Li, R., Guan, Q., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112.

30. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2024). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, *21*(1), 1–27.