INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY: APPLIED BUSINESS AND EDUCATION RESEARCH
2026, Vol. 7, No. 9, 3781 – 3790
http://dx.doi.org/10.11594/ijmaber.07.09.05

Research Article

Guided and Unguided AI Use in Assessments: Student Engagement and Satisfaction

Jelica R. Enriquez*

De La Salle-College of Saint Benilde, Philippines

Article history:
Submission 12 May 2026
Revised 31 August 2026
Accepted 23 September 2026

**Corresponding author:*
E-mail:
jelica.enriquez@benilde.edu.ph

ABSTRACT

The rapid adoption of generative artificial intelligence (GenAI) in higher education has changed how students plan, draft, revise, and complete assessment tasks. However, there is still limited evidence on how instructional guidance shapes students' experiences with these tools. This study examined whether guided use of generative AI, defined as the use of structured prompting protocols, required human-in-the-loop critical reflection, and verification workflows, produces different outcomes from unguided use. A quantitative, quasi-experimental, cross-sectional comparison was employed involving 50 undergraduate students from information technology-related programs, with 25 students in the guided AI use group and 25 in the unguided AI use group. Student engagement and learning satisfaction were measured using multi-item Likert-scale subscales administered through an online post-assessment survey, and group differences were analyzed using independent-samples t-tests after assumption checks for normality and homogeneity of variance. Results showed that the guided AI use group reported significantly higher student engagement than the unguided group. The guided group likewise obtained significantly higher learning satisfaction than the unguided group. These findings indicate that AI does not improve educational outcomes through access alone. Instead, its value depends on pedagogical structure, explicit expectations, and instructor support. The study provides quantitative evidence that guided AI use can promote more active engagement and more satisfying assessment experiences than unguided use. It also offers practical implications for designing AI-supported assessment in higher education.

Keywords: *Artificial intelligence, GenAI, Guided use, Student engagement, Student satisfaction, Unguided use*

How to cite:
Enriquez, J. R (2026). Guided and Unguided AI Use in Assessments: Student Engagement and Satisfaction. *International Journal of Multidisciplinary: Applied Business and Education Research*. *7*(9), 3781 – 3790. doi: 10.11594/ijmaber.07.09.05

Introduction

Artificial intelligence (AI) has moved from novelty to routine in many sectors, and higher education is no exception (Basch et al., 2025; Delello et al., 2025; Hooda et al., 2022). The growing use of AI in higher education has transformed the ways students interact with assessment, academic writing, and classroom learning processes. Generative AI (GenAI) tools such as ChatGPT, Gemini, and Grammarly are increasingly used by students for idea generation, summarization, drafting, revision, and feedback seeking. As a result, AI is becoming more integrated into everyday academic practice (Baidoo-Anu & Ansah, 2023; Chan & Hu, 2023; Rasul et al., 2023; Yeung, 2025). In higher education evaluations, GenAI is increasingly integrated in how students learn more (Kestin et al., 2025), and into students' planning, composing (Hostetter et al., 2024), and revising of written and oral assignments (Zhao, 2024), impacting both assessment outcomes and the processes of learning and evaluation. In this context, the key question is no longer whether students use AI, but under what instructional conditions AI supports learning rather than displacing it.

The literature so far offers two competing stories. On one side, AI-supported assessment can improve feedback quality, stimulate cognitive and metacognitive growth, and create genuinely positive learning experiences (Otto et al., 2025; Zhao, 2024). On the other, researchers warn of overreliance, passive learning, weak verification habits, academic dishonesty, and sagging motivation when students are left to use these tools without pedagogical direction (Borges et al., 2026; Kim et al., 2025; Yusuf et al., 2024). These contrasting findings indicate that the educational implications of AI are shaped less by the technology itself than by the design of the learning environment in which it is used (Zhao, 2024).

Student engagement has become an important lens for understanding these issues because it reflects students' attention, effort, motivation, and cognitive involvement in learning activities. It is often treated as a key indicator of meaningful learning. Prior studies show that GenAI can deepen emotional and cognitive engagement when it is embedded in activities with clear expectations and strong instructional support (Guo et al., 2025; Yang, 2025). However, weakly structured settings tend to produce superficial participation (Cassim & Yang, 2025).

Learning satisfaction is likewise central when evaluating AI-supported assessment, because it captures how coherent, fair, and educationally meaningful students find the experience. Satisfaction tends to increase when AI is perceived as useful (Almulla, 2024), easy to use, transparent, enjoyable (Kim et al., 2025), and aligned with task expectations (Yang, 2025). When instructional expectations are clear, AI may be perceived as a constructive academic support. However, when ethical boundaries and procedures are vague, students may experience uncertainty, distrust, and tension between originality and convenience (Almulla, 2024).

Within the broader debate on AI in education, it is important to distinguish between guided and unguided use. Guided AI use refers to assessment conditions in which instructors provide clear direction on when and how students may use AI. This direction may include guidance on developing prompts, procedures for evaluating and verifying AI-generated outputs, and ethical expectations for incorporating AI assistance into coursework. By contrast, unguided AI use refers to situations in which students are free to use AI tools with little or no structured support from the instructor (Otto et al., 2025). Existing evidence suggests that these conditions may produce different learning experiences. However, direct quantitative comparisons remain limited in authentic classroom contexts (Morgana & Poli, 2026). Despite recent developments in the use of GenAI in education, there is still limited empirical work directly comparing students who receive clear AI usage instructions and those who do not (Guo et al., 2025; Zhao, 2024).

The present study addresses this gap by comparing guided and unguided AI use in higher education assessments, with specific attention to student engagement and learning satisfaction (Almulla, 2024; Otto et al., 2025). Addressing this gap is significant for informing evidence-based guidelines on AI integration in higher education assessments and for ensuring that AI-supported evaluation practices promote meaningful engagement and satisfying learning experiences. This study has two objectives: first, to compare student engagement between guided and unguided AI use in assessments, and second, to compare student satisfaction between these two conditions. The study is situated in undergraduate classes in the Philippines, where AI use in assessment is increasingly visible but institutionally uneven. By comparing two groups exposed to similar assessment tasks but different levels of instructional guidance, the study aims to provide evidence on whether structured AI integration is associated with more favorable student outcomes, thereby contributing to emerging discussions on responsible AI adoption, digital pedagogy, and assessment design.

Materials and Methods

Research Design

The study examined a quantitative, cross-sectional, quasi-experimental two-group comparative design to examine whether explicit instructional guidance on AI use in assessment was associated with differences in students' engagement and learning satisfaction. The design was quasi-experimental because students were not individually and randomly assigned to the two conditions. One intact class served as the guided AI-use group. The unguided AI-use group was formed by randomly selecting students from two comparable classes. Because the study included two independent groups and two continuous outcome variables, independent-samples t-tests was employed as the primary inferential procedure. This test is suitable for determining whether the mean scores of two unrelated groups differ significantly on a continuous measure (Mishra et al., 2019), namely, student engagement and learning satisfaction. The grouping variable was coded as 0 = unguided AI use and 1 = guided AI use.

Participants and Sampling

The participants were 50 undergraduate students enrolled in information technology-related programs at a single higher education institution in Manila, Philippines. The guided AI use condition consisted of one intact class with 25 students, while the unguided condition drew 25 students randomly from two comparable classes. Purposive sampling was used at the class level because the selected classes were relevant to the research problem and comparable in disciplinary orientation (Etikan et al., 2016), while random selection within the unguided pool improved balance between the groups. Group sizes were deliberately kept equal, since balanced designs make the accompanying tests more robust (Cohen, 1992a; Winer et al., 1991). Across both conditions, students were predominantly first-year and second-year undergraduates, which provided reasonable comparability in academic level and day-to-day familiarity with technology.

This sampling approach may introduce selection bias. Intact classes bring pre-existing characteristics to the study, including exposure to the same instructor, an established classroom culture, and study habits developed over time. Randomly selecting students from the unguided group may reduce some differences between participants, but it cannot fully account for these pre-existing class-level influences. Accordingly, the findings of this study are interpreted as associations observed under quasi-experimental conditions, rather than as definitive evidence of causal effects.

In the guided AI condition, students received explicit instructions on when and how to use AI during assessment tasks. In practice, the guidance followed a simple protocol. Before opening any tool, students first attempted the task themselves and wrote down their own reasoning. They then used ChatGPT or Google Gemini with a brief series of prompts from the instructor. For instance, they may ask the tool to explain the reasoning behind a scenario-based inquiry before seeking an answer, or they could ask for two different answers side by side. The next phase involved students placing the AI outputs next to their own draft reasoning, noting at least one claim they wanted to check against the course materials, and identifying where the tool agreed, disputed, or contributed something new. Only after this checking step could AI-generated content be incorporated, and anything incorporated had to be restated in the student's own words. Grammarly was recommended for polishing the final text. A typical worked example: for a scenario-based question, students asked Gemini to generate two defensible answers, challenged the weaker one with a follow-up prompt, and then wrote a short justification, in their own voice, for the answer they finally chose.

In contrast, the unguided AI use condition comprised students from two similar classes. In these classes, tools such as ChatGPT, Grammarly, and QuillBot were available and permitted, but the instructor did not provide explicit prompts, examples, or step-by-step procedures for using them. All classes were 3-hour sessions taught by technology-proficient faculty using comparable technological infrastructure, including a learning management system, video conferencing tools, and institutional email, thereby reducing extraneous variability in access to digital resources across groups.

Measures

Guided AI use referred to conditions in which students received explicit instructions and structured support on when and how to use AI tools during the assessment. Unguided AI use referred to conditions in which students could use AI tools at their discretion, without such procedural guidance.

The study employed two dependent variables: student engagement and learning satisfaction. Student engagement was measured through a nine-item scale reflecting active participation, effort, motivation, interest, and cognitive involvement in learning activities involving AI tools. Learning satisfaction was measured through a six-item scale assessing students' satisfaction with AI use in the class, perceived improvement in learning experience, attainment of learning goals, and overall satisfaction with outcomes.

Instrument

The online questionnaire was developed with reference to the Technology Acceptance Model (Davis, 1989) and recent studies on GenAI-supported learning, although only the engagement and satisfaction subscales served as focal outcomes in the present analysis. The engagement and satisfaction questions were adapted from Handelsman et al. (2005) and Almufarreh (2024) respectively. The student engagement scale consisted of nine items and the learning satisfaction scale consisted of six items, all answered on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). For each subscale, item scores were averaged to form composite indices. Internal consistency reliability was examined using Cronbach's α, and an exploratory factor analysis was conducted to examine the underlying factor structure of the instrument. Faculty experts reviewed the instrument for clarity and relevance, and a small pilot group of students provided feedback for minor refinement of the instrument questions.

Data Collection Procedure

Data were collected at the end of the instructional period in April 2026 through a post-assessment online survey administered after the completion of the classroom assessment activities. The questionnaire was delivered using Google Forms, which allowed students to respond electronically using their institutional accounts. This timing allowed to capture students' reported engagement and satisfaction immediately after they experienced the AI-supported assessment in their respective conditions.

Ethical Considerations

Before accessing the survey, participants were informed about the purpose of the study, the voluntary nature of participation, and confidentiality measures, and they indicated their consent through an acknowledgment item embedded in the form.

Declaration of AI-Assisted Writing

To support transparency, the author discloses limited use of Google Gemini, Grammarly, and QuillBot to check grammar, refine syntax, and copy-edit author-written drafts. These tools were not used to generate any section of the manuscript from scratch.

The author designed the study, collected and analyzed the data, interpreted the findings, and developed all substantive arguments. The author reviewed and verified the final manuscript and accepts full responsibility for its content.

Data Analysis

Survey data from Google Forms were exported to Microsoft Excel for initial preparation and cleaning. In Excel, Likert-scale items were coded numerically using a 5-point scale, where 5 = Strongly agree, 4 = Agree, 3 = Neutral, 2 = Disagree, and 1 = Strongly disagree. Composite scores for student engagement and learning satisfaction were computed as the mean of their respective items, with higher values indicating higher engagement and higher satisfaction.

Subsequent analyses were conducted using Jamovi. Descriptive statistics, including frequency, mean, median, standard deviation, minimum, maximum, skewness, kurtosis, and Shapiro–Wilk statistics, were generated for student engagement and learning satisfaction separately for the guided and unguided groups. These descriptives confirmed that each group contained 25 respondents and that the guided group obtained higher average scores on both outcomes than the unguided group. Assumption checks included tests of normality (Shapiro–Wilk, supplemented by additional normality tests and visual inspection) and homogeneity of variances (Levene's test and variance ratio tests) for each outcome by group.

Independent-samples t-tests were conducted in Jamovi to compare the mean scores of the guided and unguided groups on student engagement and learning satisfaction. The assumption checks indicated that variances were homogeneous and that deviations from normality were moderate, supporting the use of the Student independent-samples t-test as the primary inferential method.

Result and Discussion

Descriptive Results

The guided AI use group obtained higher mean scores than the unguided group on both outcome variables. For student engagement, the guided group (n = 25) had a mean of 4.27 SD = 0.56, whereas the unguided group (n = 25) had a mean of 3.84 SD = 0.57. For learning satisfaction, the guided group had a mean of 4.17 SD = 0.69, while the unguided group had a mean of 3.72 SD = 0.65. These values already suggested that students who received explicit AI guidance reported more favorable experiences than those who used AI without structured support. The descriptives are presented below in Table 1.

Table 1. Descriptive statistics for student engagement and learning satisfaction by group

	Group	N	Mean	Median	SD	SE
Student Engagement	0	25	4.27	4.11	0.555	0.111
	1	25	3.84	3.89	0.571	0.114
Learning Satisfaction	0	25	4.17	4	0.693	0.139
	1	25	3.72	3.83	0.645	0.129

Visual and numerical distribution checks indicated relatively symmetric distributions without severe outliers for both variables. The means and medianswithin groups were closely aligned, suggesting limited skewness in the composite scores. The Quantile-Quantile (Q-Q) plots for Student Engagement and Learning Satisfaction are presented below in Figure 1.

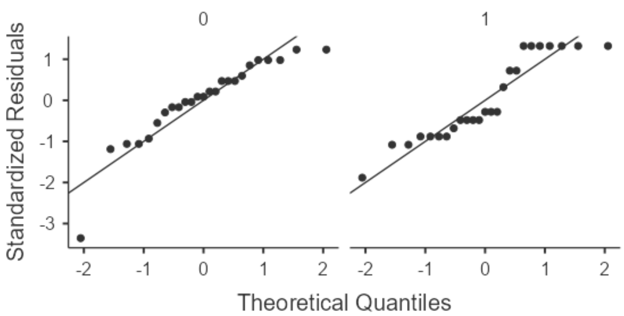

Figure 1. Quantile-quantile plots for Student Engagement and Learning Satisfaction

Homogeneity of variance assumptions were also met, as Levene's tests were non-significant for student engagement $F(1,48) = 0.57$, $p = .454$ and learning satisfaction $F(1,48) = 1.43$, $p = .238$. Although the Shapiro–Wilk test for learning satisfaction was statistically significant $W = 0.94$, $p = .015$, the combined assumption checks and balanced group sizes supported the robustness of the independent-samples t-test for the present analysis. The test results of homogeneity of variance are presented below in Table 2.

Table 2. Homogeneity of Variances

		F	df	df2	p
Student Engagement	Levene's	0.571	1	48	0.454
	Variance ratio	0.943	24	24	0.887
Learning Satisfaction	Levene's	1.427	1	48	0.238
	Variance ratio	1.154	24	24	0.728

Given the continuous outcomes, the balanced two-group design, and the assumption checks, the independent-samples t-test was retained as the main inferential approach and was deemed appropriate for addressing the research objectives (Ghasemi & Zahediasl, 2012; Sawilowsky & Blair, 1992).

Inferential Results

Considering the level of measurement of the outcomes, the balanced two-group design, and the assumption checks, the student independent-samples t-test (Ghasemi & Zahediasl, 2012; Sawilowsky & Blair, 1992) was retained as the main inferential approach, and deemed appropriate for addressing the study's central research questions about mean differences in engagement and satisfaction between guided AI use and unguided AI use in higher education assessment.

Independent-samples t-tests revealed statistically significant differences between the guided and unguided groups for both student engagement and learning satisfaction. For student engagement, the guided AI use group ($M = 4.27$, $SD = 0.56$, $n = 25$) scored significantly higher than the unguided AI use group ($M = 3.84$, $SD = 0.57$, $n = 25$), $t(48) = 2.65$, $p = .011$, mean difference = 0.42, 95% CI [0.10, 0.74], $d = 0.75$. This finding suggests that students who received structured AI instructions showed higher engagement than those who used AI without explicit guidance.

A similar pattern was observed for learning satisfaction. Students in the guided AI use group ($M = 4.17$, $SD = 0.69$, $n = 25$) reported significantly higher satisfaction than students in the unguided AI use group ($M = 3.72$, $SD = 0.65$, $n = 25$), $t(48) = 2.39$, $p = .021$, mean difference = 0.45, 95% CI [0.07, 0.83], $d = 0.68$. This result indicates that explicit instructional guidance on AI use was associated with higher satisfaction with the learning experience.

In terms of effect size, the observed Cohen's d values (Cohen, 1992b) indicate that both differences were not only statistically significant, but also practically meaningful. In practical terms, the guided condition produced moderate improvements in both engagement and satisfaction. The effect size for student engagement ($d = 0.75$) approaches a moderate-to-large range, while the effect size for learning satisfaction ($d = 0.68$) falls within the moderate range, indicating that the guided AI use condition had a meaningful positive association with both outcomes. These results suggest that structured AI guidance contributes positively to students' engagement and satisfaction in higher education assessment contexts. The independent samples t-test are presented below in Table 3.

Table 3. Independent-Samples T-Test Results for Guided Versus Unguided AI Use

		Statistic	df	p	Mean difference	SE difference	95% Confidence Interval Lower	95% Confidence Interval Upper		Effect Size
Student Engagement	Student's t	2.65	48	0.011	0.422	0.159	0.1021	0.742	Cohen's d	0.75
Learning Satisfaction	Student's t	2.39	48	0.021	0.453	0.189	0.0726	0.834	Cohen's d	0.67

Note. Guided group n = 25; unguided group n = 25.

The outcomes of this research reveal that guided use of AI in assessment differs from unguided use of AI in terms of student involvement and learning satisfaction among undergraduate students. Students who received explicit guidance on how to use AI tools reported significantly higher engagement and significantly higher learning satisfaction than students who used AI without structured instructional support. These findings give empirical support for the main premise of this research that the educational impacts of AI in higher education are shaped not simply by access to the technology itself, but by the pedagogical conditions under which that technology is introduced and utilized.

For student engagement, the guided group obtained significantly higher scores than the unguided group, suggesting that structured AI guidance may promote more active, effortful, and cognitively focused participation in assessment-related learning. This finding is consistent with prior studies suggesting that generative AI can enhance emotional and cognitive engagement when it is embedded in contexts characterized by meaningful support and instructional challenge (Guo et al., 2025; Yang, 2025). In particular, Guo et al. (2025) argued that AI has no inherent educational value in the absence of the instructional design that surrounds it, and the current results are consistent with the position in that students were more engaged when AI use was guided by explicit instructor direction rather than left to autonomous experimentation. The result also resonates with Yang's (2025) claim that AI-supported learning can strengthen motivation and cognitive engagement when students are guided toward purposeful interaction with the technology, rather than using it in a convenience-driven way.

Practical Implications

The findings carry practical implications for higher education instructors and academic leaders. First, AI-supported assessment should be accompanied by explicit guidance on when AI may be used, how outputs should be evaluated, and how generated content can be incorporated ethically into student work. Second, institutions should consider developing standardized AI-use guidelines or classroom protocols that reduce ambiguity and promote responsible student behavior during assessments. Third, digital pedagogy initiatives should move beyond tool familiarization and focus instead on structured academic uses of AI that preserve cognitive demand and student agency.

For curriculum designers, AI guidance should be built into the curriculum rather than left to individual classroom decisions. Short AI-literacy components, including prompt development, output verification, and ethical use, can be incorporated into assessment criteria and rubrics so that students receive consistent guidance across course sections. Assessment criteria should also value the process behind AI-supported work, including how students verify, evaluate, and justify their use of AI, rather than rewarding only a polished final output that AI can easily create.

More broadly, the findings suggest that successful AI integration requires more than access to digital tools. Institutions need clear policies, practical guidance, and instructional processes that preserve students' cognitive engagement and agency. Providing unguided access may expand availability, but structured use is more likely to support positive learning experiences.

Limitations

This study was conducted with students from a single higher education institution in Manila, Philippines. Therefore, the findings may not generalize to other institutional, cultural, or disciplinary settings. The sample was drawn from IT programs, and future research should include students in humanities and business disciplines to examine whether disciplinary context affects the value of guided AI use.

In addition, the study used intact classes rather than fully randomized assignment. Although efforts were made to improve group comparability, pre-existing differences between classes may still have influenced the results. The study also relied on self-reported measures of engagement and learning satisfaction rather than behavioral indicators or academic performance outcomes.

Finally, the intervention covered only one two-week assessment cycle. Longer-term studies across multiple assessments are necessary to determine whether the observed benefits of guided AI use persist over time.

Conclusion

This study provides quantitative evidence that guided AI use in higher education assessments is associated with higher student engagement and higher learning satisfaction than unguided AI use. Students who received explicit instruction on how to use AI during assessment reported more positive learning experiences than students who were merely allowed to use AI without structured support. These findings indicate that AI becomes more educationally valuable when embedded within clear pedagogical expectations, ethical boundaries, and purposeful classroom procedures. More broadly, the study suggests that the success of AI-supported assessment depends not only on the availability of generative tools but also on the quality of instructional design, including how instructors frame tasks, model AI use, and set ethical expectations. Future research may extend this work by including academic performance outcomes, comparing multiple levels of guidance, and using mixed-method or longitudinal approaches to capture how students' experiences with guided AI use develop over time.

The main findings (novelty) and why they are important and useful:

This study shifts empirically from examining whether students use Generative AI (GenAI) to examining how instructional design determines the quality of GenAI use. This study offers quantitative evidence that provides the results of guided AI use from unguided AI usage in classrooms, while previous research mostly remained concentrated on tool-specific perspectives. This study provides significant implications for educators, academic leaders, and researchers. This study supports the role of educators as pedagogical mediators in the AI era for structured development. Likewise, this study offers data-driven justification for moving beyond "ban or allow AI policies" in educational institutions. It emphasizes that successful AI integration is a matter of academic governance. Finally, this study establishes a baseline for future comparative studies involving academic performance and longitudinal effects of structured GenAI use.

References

Almufarreh, A. (2024). Determinants of Students' Satisfaction with AI Tools in Education: A PLS-SEM-ANN Approach. *Sustainability*, *16*(13), 5354. https://doi.org/10.3390/su16135354

Almulla, M. A. (2024). Investigating influencing factors of learning satisfaction in AI ChatGPT for research: University students perspective. *Heliyon*, *10*(11). https://doi.org/10.1016/j.heliyon.2024.e32220

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.2139/ssrn.4337484

Basch, C., Hillyer, G., Gold, B., & Yousaf, H. (2025). Understanding artificial intelligence knowledge and usage among college students: Insights from a survey on classroom, coursework, and personal applications. *Review of Artificial Intelligence in Education*, *6*(i). https://doi.org/10.37497/rev.artif.intell.educ.v6ii.37

Borges, I., Sebastião, C., Caridade, C., & Pereira, V. (2026). From ChatGPT to classroom learning: Exploring the role of teacher mediation in AI-supported education. *Australasian Journal of Educational Technology*, *42*(3). https://ajet.org.au/index.php/AJET/article/view/11056

Cassim, F. A. K., & Yang. (2025, December 1). GAI pedagogical agents and teacher-guided prompt use in EFL debate training. *International Conference on Computers in Education*. ICCE 2025: The 33rd International Conference on Computers in Education.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Cohen, J. (1992a). A power primer (Pt. 155-159). *Psychological Bulletin*, *112*(1). https://doi.org/10.1037/0033-2909.112.1.155

Cohen, J. (1992b). Statistical Power Analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. JSTOR.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Delello, J. A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., & De Giuseppe, T. (2025). AI in the classroom: Insights from educators on ssage, challenges, and mental health. *Education Sciences*, *15*(2), 113. https://doi.org/10.3390/educsci15020113

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1). https://doi.org/10.11648/j.ajtas.20160501.11

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Guo, F., Zhang, L., Shi, T., & Coates, H. (2025). Whether and when could generative AI improve college student learning engagement? *Behavioral Sciences*, *15*(8), 1011. https://doi.org/10.3390/bs15081011

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A Measure of College Student Course Engagement. *The Journal of Educational Research*, *98*(3), 184–192. https://doi.org/10.3200/JOER.98.3.184-192

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*(1), 5215722. https://doi.org/10.1155/2022/5215722

Hostetter, A. B., Call, N., Frazier, G., James, T., Linnertz, C., Nestle, E., & Tucci, M. (2024). Student and faculty perceptions of generative artificial intelligence in student writing. *Teaching of Psychology*, *52*(3), 319–329. https://doi.org/10.1177/00986283241279401

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kim, J., Klopfer, M., Grohs, J. R., Eldardiry, H., Weichert, J., Cox, L. A., & Pike, D. (2025). Examining faculty and student perceptions of generative AI in university courses. *Innovative Higher Education*,

50(4), 1281–1313. https://doi.org/10.1007/s10755-024-09774-w
Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia*, *22*(4), 407–411. https://doi.org/10.4103/aca.ACA_94_19
Morgana, V., & Poli, F. (2026). Guided and unguided GenAI tasks for learning academic spoken collocations. *Journal of English for Academic Purposes*, *80*, 101638. https://doi.org/10.1016/j.jeap.2026.101638
Otto, S., Ejsing-Duun, S., & Lindsay, E. (2025). Disruptive tensions and emerging practices: An exploratory inquiry into student perspectives on generative artificial intelligence in a problem-based learning environment. *Education and Information Technologies*, *30*(13), 19111–19140. https://doi.org/10.1007/s10639-025-13533-5
Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. de O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heatcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching*, *6*(1), 41–56. https://doi.org/10.37074/jalt.2023.6.1.29
Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departures from population normality. *Psychological Bulletin*, *111*(2), 352–360. https://doi.org/10.1037/0033-2909.111.2.352
Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). McGraw-Hill.
Yang, H. (2025). Harnessing generative AI: Exploring its impact on cognitive engagement, emotional engagement, learning retention, reward sensitivity, and motivation through reinforcement theory. *Learning and Motivation*, *90*, 102136. https://doi.org/https://doi.org/10.1016/j.lmot.2025.102136
Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing*, *68*, 101203. https://doi.org/10.1016/j.jslw.2025.101203
Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, *21*(1), 21. https://doi.org/10.1186/s41239-024-00453-6
Zhao, C. (2024). AI-assisted assessment in higher education: A systematic review. *Journal of Educational Technology and Innovation*, *6*(4), 39–58. https://doi.org/10.61414/jeti.v6i4.209