Artigo original http://dx.doi.org/10.18623/rvd.v23.5341

EXAMINING THE EFFECTS OF AI-DRIVEN LEARNING ANALYTICS ON PERSONALIZED FEEDBACK IN BLENDED HIGHER EDUCATION COURSES: EVIDENCE FROM NANJING NORMAL UNIVERSITY, CHINA

ANÁLISE DOS EFEITOS DA ANÁLISE DE APRENDIZAGEM BASEADA EM IA NO FEEDBACK PERSONALIZADO EM CURSOS MISTOS DE ENSINO SUPERIOR: EVIDÊNCIAS DA UNIVERSIDADE NORMAL DE NANJING, CHINA

Article received on: 11/4/2025
Article accepted on: 2/6/2026

Xu Min*
*Universitas Pendidikan Ganesha, Indonesia
xumin@uiumail.com

Ni Nyoman Padmadewi*
*Universitas Pendidikan Ganesha, Indonesia
xumin@uiumail.com

Luh Putu Artini*
*Universitas Pendidikan Ganesha, Indonesia
xumin@uiumail.com

I Gede Budasi*
*Universitas Pendidikan Ganesha, Indonesia
xumin@uiumail.com

Zhang Tao*
*Universitas Pendidikan Ganesha, Indonesia
xumin@uiumail.com

The authors declare that there is no conflict of interest

Abstract
The swift integration of Artificial Intelligence (AI) and Learning Analytics (LA) in higher education has opened the new opportunities to provide personalized feedback in scale blended learning. Nonetheless, the currently available literature is mostly focused on the efficiency of technologies and does not focus much on how students perceive and respond to AI-generated feedback. This paper examines the impacts of learning analytics powered by AI in the context of personalized feedback in blended classes of English as a Foreign Language (EFL) at Nanjing Normal University, China. The study applies a convergent mixed-methods design, which involves the quantitative analysis of LMS log data, surveys, and structural equation modeling along with qualitative analysis of feedback threads and semi-structured interviews with students and instructors. The article investigates the effects of AI-generated feedback on student engagement and feedback uptake, and academic performance, with the mediating impact of feedback literacy and a moderating effect of

Resumo
A rápida integração da Inteligência Artificial (IA) e da Análise de Aprendizagem (LA) no ensino superior abriu novas oportunidades para fornecer feedback personalizado em aprendizagem combinada em escala. No entanto, a literatura atualmente disponível concentra-se principalmente na eficiência das tecnologias e não se concentra muito em como os alunos percebem e respondem ao feedback gerado pela IA. Este artigo examina os impactos da análise de aprendizagem alimentada por IA no contexto do feedback personalizado em aulas combinadas de Inglês como Língua Estrangeira (EFL) na Universidade Normal de Nanjing, China. O estudo aplica um projeto convergente de métodos mistos, que envolve a análise quantitativa de dados de log do LMS, pesquisas e modelagem de equações estruturais, juntamente com a análise qualitativa de tópicos de feedback e entrevistas semiestruturadas com alunos e instrutores. O artigo investiga os efeitos do feedback gerado por IA no envolvimento dos alunos, na aceitação do feedback e no

instructor mediation. The findings indicate that adaptive and personalized AI feedback is a significant way of stimulating behavioral engagement, increasing the use of feedback, and leading to improved learning results. Qualitative findings also indicate that the efficacy of AI feedback is determined by its specificity, actionability, and pedagogical relevance besides the ability of students to interpret feedback and the role of instructors in contextualizing it. The study, in general, points to the idea of AI-based feedback as a human-based pedagogical instrument instead of a strictly technological one, with valuable implications to the scalable and equitable application of AI in educational institutions.

Keywords: Artificial Intelligence (AI). Learning Analytics (LA). Personalized Feedback. Blended Learning. Student Engagement. Feedback Literacy. English as a Foreign Language (EFL). Higher Education.

desempenho acadêmico, com o impacto mediador da alfabetização em feedback e um efeito moderador da mediação do instrutor. Os resultados indicam que o feedback adaptativo e personalizado da IA é uma forma significativa de estimular o envolvimento comportamental, aumentar o uso do feedback e levar a melhores resultados de aprendizagem. Os resultados qualitativos também indicam que a eficácia do feedback da IA é determinada por sua especificidade, aplicabilidade e relevância pedagógica, além da capacidade dos alunos de interpretar o feedback e do papel dos instrutores em contextualizá-lo. O estudo, em geral, aponta para a ideia do feedback baseado em IA como um instrumento pedagógico baseado no ser humano, em vez de estritamente tecnológico, com implicações valiosas para a aplicação escalável e equitativa da IA em instituições educacionais.

Palavras-chave: *Inteligência Artificial (IA). Análise de Aprendizagem (LA). Feedback Personalizado. Aprendizagem Combinada. Envolvimento dos Alunos. Alfabetização em Feedback. Inglês como Língua Estrangeira (EFL). Ensino Superior.*

1 INTRODUCTION

Higher education has undergone a swift process of digitization, which has generated a vast amount of information about students learning behaviors, especially in large-scale courses that incorporate blended elements. This vast influx of information has precipitated the convergence of Artificial Intelligence (AI) and Learning Analytics (LA) to deliver personalized feedback, which has been much desired in pedagogy to meet the needs of diverse learners. AI-driven analytics provide a scalable solutions to the problem in high-enrollment settings, such as English as a Foreign Language (EFL) education, where instructor feedback is logistically infeasible on an individual basis and where timely and data-driven feedback on the learner is needed to facilitate language acquisition.

Nonetheless, recent discussion of AI in education is typically marked by the technological determinism which prioritizes algorithmic accuracy over pedagogical value. The current literature is largely preoccupied with the technical aspects of AI to produce feedback, and little is done with the reception, interpretation, and actions of the feedback by students in a real-world learning ecosystem (Yildiz *et al*., 2025; Niemi, H. 2024; Rodrigues *et al*., 2025; Grba, D. 2022) . This is specifically true in the area of language learning, as feedback cannot be understood as a simple correction of mistakes, but a complex communicative process that has to be contextualized and understood mutually. The most crucial functions of the instructor as a mediator and the feedback literacy of the student, their ability to process and generate feedback, have not been sufficiently examined in the AI-integrated setting.

This research paper fills this gap as it does not focus on a mere comparison of AI and no-AI states. It examines the subtle effects of AI-informed learning analytics on individual feedback in a large-scale blended course in EFL learning at Nanjing normal university in China. Our approach is holistic and we are not just looking at the immediate consequences of AI generated feedback on student engagement and performance but also at how the instructional design, and instructor intervention and student feedback literacy mediate and moderately influence student engagement and performance. Through an effective mixed-methods design that integrates a quantitative modeling method with a qualitative analysis of the results of students and instructors, this study aims at learning how and in what circumstances AI can successfully complement the feedback process. The results are expected to add a human-centered paradigm of introducing analytics in a manner that is scalable and fair and shape pedagogical practice in EFL and the field of higher education in general.

2 BACKGROUND

The current work lies in between two major trends in the field of Chinese higher education as national scale implementation of blended learning in the teaching of English as a Foreign Language (EFL) and the intensive implementation of Artificial Intelligence (AI) to respond to the pedagogical challenges that the said change implies. This twofold context is crucial because it will base the research on a real institutional reality, as opposed to a theorized AI implementation.

A compulsory course in China known as College English is required by millions of non-English majors undergraduate students, which has generated an enormous demand that is far beyond the capacity to serve. The recent scholarship captures the response of the Chinese universities to this scale challenge through the introduction of blended learning models commonly used in Chinese universities to combine Learning Management Systems (LMS), mobile applications and collaborative learning models. According to Su, Sazalli, and Miskam (2024), these models will cater to various needs of learners and handle the high population of students, which means that the technology-based delivery is not only innovative but it is structurally required in the modern context of education.

In this hybrid ecosystem, AI-assisted instruction has become one of the potential solutions to the problem of at-scale-personalized language education. A more recent quasi experimental research by Cao and Phongsatha (2025) regarding the use of a Foreign Language Intelligent Teaching (FLIT) platform in China is a strong indication of the potential of AI. Their results show that AI-enhanced blended learning provides considerable gains in reading, listening, writing, and speaking skills in Business English, and, at the same time, it brings positive gains in the cognitive and behavioral engagement levels in contrast to the conventional teaching approach.

Taken together, all this emerging research supports the fact that AI-based Learning Analytics promise much in the Chinese EFL settings. These studies, however, also indirectly point at a very important caveat: technological potential is not necessarily pedagogically effective. The mentioned positive results will depend on the considerations of feedback design and, most importantly, the ability of learners to process and respond to AI-generated information. This highlights the need to explore not only the possibility of AI creating feedback, but how the students interact with it, and this is precisely what the current study fills in by studying the mediating effects of feedback literacy and instructor intervention in a natural setting of a Chinese university.

3 CONCEPTUAL FRAMEWORK

Figure 1 illustrates the conceptual framework that will be used in the current study. The essence of the model is that AI-assisted learning analytics (LA)-based personalized feedback with its inherent features of timeliness, specificity, transparency, and actionability is the key antecedent affecting student engagement on behavioral, cognitive, and emotional levels. With this process, it is likely to influence other learning outcomes like assessment and grading work. This cycle is indicative of how the analytics systems have been established to recognize diagnostic clues and prescribe interventions, which once identified and implemented by the learners will encourage behavior change and eventually lead to better academic outcomes (Ancess, J. 2000; Boekaerts, M. 1992; Hodge, B. 2014; Hungerford, H. R., & Volk, T. L. 1990; Luo *et al*., 2025; Warren *et al*., 2006). The solid arrow in the model symbolizes the relationship between feedback design and engagement indicators, which are measured using LMS log data and course interaction records (Kittur, J *et al*., 2022; Wang, F. H. 2017; Henrie *et al*., 2018; Park, Y., & Jo, I. H. 2017); therefore, the direct direction of the feedback created by AI-driven LA to student engagement (Kaliisa *et al*., 2024).

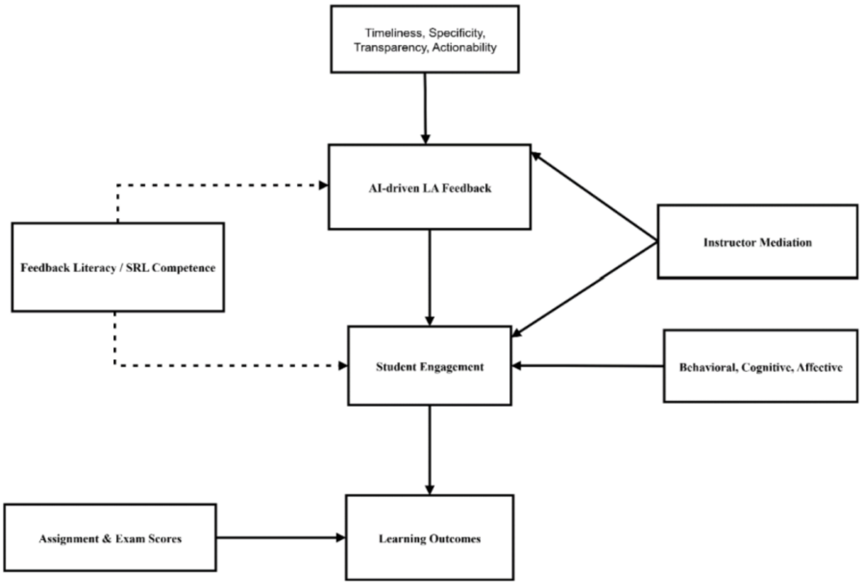
Furthermore, feedback and self-regulated learning skills are integrated into the framework as a mediating variable between feedback and engagement, like a dotted arrow pointing to the main causal relationship. This suggests that the effectiveness and direction of analytical feedback depend on students' ability to interpret and evaluate that feedback, and to act on it. Empirically, it has been shown that a greater degree of feedback literacy results in a higher perceived usefulness and a better acceptance of the outputs of informative analytics, and a lesser degree results in cognitive overload or misinterpretation of feedback (Tepeç *et al*, 2024; Weidlich *et al*, 2025). The moderating line that is visually emphasized in the model is important to note because it is true that feedback literacy does not only introduce a new effect to the model but also changes the effectiveness of feedback to induce engagement and behavioral change. The other contextual design aspect used in the framework is the Instructor Mediation which influences the AI-based outputs as well as helping students to understand the analytic suggestions. Previously published experimental studies indicate that the use of analytical and contextualized reflective messages, supports, or stimuli provided by teachers, rather than automated feedback, will significantly increase student engagement and academic achievement, especially among students with less developed self-regulated learning (SRL) (Zhou Q *et al*., 2022).

In the diagram, instructor mediation is shown as an applicable relational path (a second arrow leading to student engagement is located on the feedback engagement relationship). This is both indicative of its dual role, the first to enhance the pedagogical alignment and intelligibility of analytics outputs, and the second, to enable students whose feedback literacy is limited, to make sense of the recommendations and take action. Finally, the Student Engagement to Learning Outcomes pathway demonstrates the mediating nature of engagement and SRL behaviors that are expected. When feedback on analytics can lead to increased timely, focused engagement, i.e., increased time on task, work revision, and specific practice, it is likely to produce better results in assignments and exams (Uzun Y. *et al*., 2025; Lim L. *et al*., 2021;).

Based on this, the framework will give rise to a number of testable hypotheses: (H1) more specific and delivered within a small time-span AI-driven feedback will have a highly positive effect on engagement; (H2) moderated by feedback literacy, the former will amplify the effect of feedback among students with a higher literacy level; and (H3) the engagement level will reflect the effect of feedback on academic performance. Together, they form a conceptual model that is both theoretically sound and practically relevant to conducting empirical research on the situation with blended higher education (Luo *et al*., 2025; Tepeç *et al*., 2024). The suggested conceptual framework is shown in Figure 1: The research paper focuses on the application of an AI-based Learning Analytics system, as shown by Figure 1 to blended courses at Nanjing normal University. The schema represents the visualization of the interconnection between major elements: It examines how the data on Student Engagement (including behavioral, cognitive, and affective areas) along with Assignment and Exam Scores are processed using AI systems (Weidlich *et al*., 2025). AI processes data and provides feedback that exhibits the four essential qualities of speed, specificity, transparency, and usability, as the model demonstrates. The study evaluates the direct effects of such automated feedback on development of two essential student competencies Feedback Literacy and Self-Regulated Learning (SRL) Competence. The most significant question to be raised is how this cyclical process can be used to produce Personalized Feedback and later enhance Learning Outcomes. Importantly, the study places this contextually in the Chinese higher education through the mediating role of instructors, which is one of the linkages in Figure 1. It discusses how educators at Nanjing normal university make sense of AI-generated insights to improve their pedagogies so that the technology does not replace the human factor but improves it in the blended learning setting.

Figure 1

Conceptual Framework



4 RESEARCH METHODOLOGY

4.1 Research design and rationale

The research methodology used in this study is a convergent mixed-methods approach, meaning it combines quantitative and qualitative elements to capture both the objective effects of feedback from AI-based learning analytics and the interpretive processes by which students and teachers understand and respond to this feedback. The use of a mixed-methods approach is justified by the fact that the research questions allow (a) determine the causal pathways and the magnitude of their effects, which is best measured quantitatively and addressed with the help of statistical models, and (b) explain meaning-making in contexts, which is best answered through qualitative research (Creswell and Plano Clark, 2018). This convergent method is highly suggested in the current LA studies (Yu, H., & Bertsekas, D. P. 2009; Carlson, K. D., & Herdman, A. O. 2012; Alrawashdeh *et al*., 2021; Klahr, D., & Simon, H. A. 1999; Losos, J. B. 2011), since it enables the log-data and quasi-experimental contrasts to be triangulated with the evidence of the interview and survey, thus unpacking the black-boxes of mechanisms. Recent systematic reviews also emphasize that design features and human mediation are as important as algorithmic accuracy, thus validating the need for an integrated methodology capable of both testing hypotheses and interpreting observed trends (Banihashem *et al*., 2022; Luo *et al*., 2025). Therefore, this study synthesizes quasi-experimental comparisons, analysis of online learning platform logs, validated survey measures, automated coding of comment content, and semi-structured interviews to develop a coherent, multi-source explanation of how data-driven commenting functions in undergraduate hybrid English courses at a Chinese public university.

4.2 Study setting and participants

The focus of this study is Nanjing Normal University, a large Chinese public institution where English as a Foreign Language (EFL) is taught using a hybrid model: weekly in-person sessions are supplemented by substantial online work organized on a private learning platform. The university recently integrated an AI-powered analytics module into its platform that automatically generates feedback reports and personalized micro-recommendations for students. The pedagogical appropriateness of this environment is that Chinese public universities often receive high numbers of students into College English courses (Wang *et al*., 2021; Yan, J., & Huizhong, Y. 2006; Hu, G. 2002), generating the desire to provide such a high volume of feedback (personalized) and, as a result, the need to scale it (LA deployment) is achievable due to the presence of digital infrastructure (Su, Sazalli, and Miskam, 2024; Cao and Phongsatha, 2025).

The participants will be about 400 undergraduate students who study various courses. The quasi-experimental design with a section-level will be used: the sections under treatment will be provided with additional AI-based feedback with a focus on the short latency, high specificity, clear explanation and recommendations about the actions; the sections under the control will be given the conventional instructor feedback and standard LMS notifications. The students' basic demographic and academic characteristics, such as their prior English proficiency, will be documented to facilitate covariate adjustment and propensity score matching, thereby minimizing selection bias (Roman, M. *et al*., 2025; Benhadj, Y. *et al*., 2021). To provide more depth and context to the quantitative results, a targeted qualitative subsample of 2,530 students with high and low feedback competence profiles and 68 teachers will be invited to participate in semi-structured interviews.

The following figure 2 shows research methodology used in this study in the form of a diagram: As shown in the figure, the mixed-methods research design is convergent in that it combines quantitative and qualitative strands. The quantitative component will include analysis of LMS login logs and coding of comments using artificial intelligence, a study of comment mastery and self-regulated learning (SL), and a quasi-experimental comparison between groups. PLS-SEM will be used to test these data through mediation and moderation. The qualitative component will include semi-structured interviews with students and teachers, as well as document and comment analysis and thematic analysis. Both components will be conducted concurrently and given equal priority. There is an integration of findings based on joint display and meta-inferences, whereby through such a method complete interpretation of findings is achieved through triangulation of statistical relationships and through the rich contextual information on feedback processes and learning outcomes.

Figure 2

Research Methodology

5 DATA ANALYSIS

5.1 Demographic characteristics

Table 1

Demographic Characteristics of Students and Instructors

Characteristic	Category	Treatment (n = 200)	Control (n = 200)	Total (N = 400)
Gender	Male	90 (45.0%)	90 (45.0%)	180 (45.0%)
	Female	110 (55.0%)	110 (55.0%)	220 (55.0%)
Age (Years)	Mean (SD)	19.4 (1.2)	19.6 (1.3)	19.5 (1.2)
Year of Study	First Year	120 (60.0%)	120 (60.0%)	240 (60.0%)
	Second Year	80 (40.0%)	80 (40.0%)	160 (40.0%)
Academic Major	Business	60 (30.0%)	60 (30.0%)	120 (30.0%)
	Education	50 (25.0%)	50 (25.0%)	100 (25.0%)
	Engineering	90 (45.0%)	90 (45.0%)	180 (45.0%)
Prior English Proficiency	Mean (SD)	75.2 (9.8)	74.8 (10.2)	75.0 (10.0)
Instructors	N	4	3	7

Table 1 illustrates the demographic features of the study subjects who are students and instructors respectively in the treatment and control groups (N = 400). The groups were of 200 students each, and the groups were balanced to compare to each other. In terms of gender distribution, the sample had 180 male students (45.0%), and 220 female students (55.0%). Gender balance between the conditions was maintained as the proportions were the same in both the treatment and control group with 90 males (45.0) and 110 females (55.0) respectively. The average age of the respondents was similar. The mean age of the students in the treatment group was 19.4 years (SD = 1.2), with a mean age of the students in the control group standing at 19.6 years (SD = 1.3). The average age was 19.5 years (SD = 1.2) indicating that there is not much age difference between groups.

Regarding the year of study, 240 students (60.0% were first-year students, and 160 students (40.0% were second-year students. Distribution was once again balanced in the groups with 120 first-year (60.0%) and 80 second-year students in the treatment and control groups respectively.

There was also even distribution of academic majors. The number of business students (30.0%) was 120, Education students (25.0%) were 100 and Engineering students (45.0%) were 180. The experiment conditions were the same in proportional representation to each of the major.

In area of previous English proficiency, the treatment group was reported to have mean of 75.2 (SD = 9.8) whereas the control group had a mean of 74.8 (SD = 10.2). The general proficiency mean of 75.0 (SD = 10.0) showed that the groups had an equal level of language proficiency in the baseline. Lastly, there were seven instructors who were involved in the study, four belonging to the treatment group and three to the control group.

In general, the demographic variables indicate the treatment and control groups were balanced in most important background variables which favour the internal validity of comparative analyses in the future.

5.2 Quantitative analysis

The platforms (LMS/VLE, LXP, and the AI feedback system) will record all the events, which will be time-stamped and related to individual students, courses, and weeks. After identification, raw log data will be processed using automated system pings and duplication, and aggregated into weekly and course-level measures that are frequently used in recent studies of learning analytics (LA) on engagement and feedback usage. The measures are frequency of logins, number of sessions, number of resources accessed, number of posts, assignment timeliness rates, time-on-task, variability (averageness of the click entropy), and feedback uptake measures, such as amount of posts revised by a user post-comment, number of requests to clarify a post, or number of posts responded to by a peer in response to instructor/AI feedback. The choice of indicators and aggregation processes is consistent with the modern research of LA and dashboard design that focuses on interpretable and behaviorally based measures that enable informed decisions and interventions in a timely manner (Bergdahl *et al*., 2024).

Preprocessing procedures, such as tokenization, the removal of English and Chinese stop words where possible, and lemmatization, will be applied to the AI-generated, teacher-written text feedback. A hybrid analytical framework will then be used: (i) topic modeling with LDA/NMF to ensure transparency and comparability; (ii) lexical embeddings based on transformers (e.g., sentence-BERT) to measure semantic similarity; (iii) supervised classification of feedback functions (e.g. directive, facilitative, elaborative explanation or metacognitive prompting) using a hand-annotated reference dataset. Comparative analysis of topic modeling and transformer-based methods will be used to make the methodological decisions and map out the process of human validation of outputs (Romero *et al*., 2024; Sheils *et al*., 2024).

Personalization will be measured using three components, including: (a) lexical-semantic alignment between feedback and the error profile of the individual student (measured by calculating cosine similarity between feedback and the error vectors of the individual student), (b) specificity, determined by the presence of task-related and evidence-based cues; and (c) adaptability, measured by increasing difficulty and depth of scaffolding, based on recent work on AI feedback establishing a correlation between adaptive and non-adaptive feedback. These measures will be standardized (z-scores) and, where appropriate, aggregated into a higher-level construct of feedback personalization for structural modeling (Bergdahl *et al*., 2024). Table 2 presents the descriptive statistics of the questionnaire and the collected data; all analyses were performed using SPSS.

Table 2

Descriptive Statistics and Correlations among Study Variables (N = 400)

Variable	1	2	3	4	5
1. AI Feedback Quality	1.00				
2. Feedback Literacy	0.35**	1.00			
3. Behavioral Engagement	0.45**	0.40**	1.00		
4. Assignment Scores	0.28*	0.32*	0.47**	1.00	
5. Exam Scores	0.25*	0.30*	0.45**	0.78**	1.00

Note: N = 400. *p < 0.05, **p < 0.01. Correlations computed using Pearson’s r, with pairwise deletion for missing data.

The comparison between the treatment and control groups in terms of AI feedback quality, feedback comprehension, behavioral engagement, homework results, and exam results is represented in Figure 3, as shown below:

Figure 3

Comparison of treatment and control groups regarding AI feedback quality, feedback comprehension, behavioral engagement, assignment and exam scores.

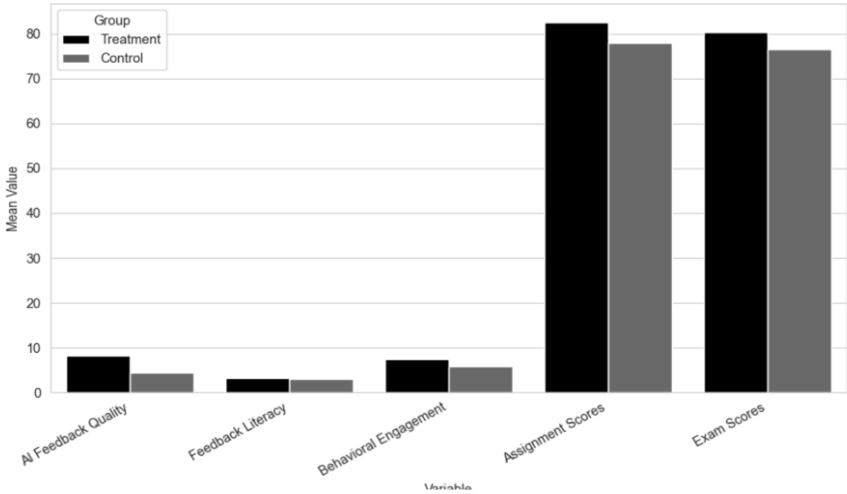


Figure 3 provides a comparative study of treatment and control groups on five major variables. The treatment group show significantly better AI feedback quality and behavioral attendance as well as moderately better assignment and exam scores. There are minimum differences in the feedback literacy. In general, the number indicates that there are positive academic and engagement outcomes that AI-enhanced feedback can have.

Figure 4

Mean comparisons between treatment and control groups with standard deviation error bars.

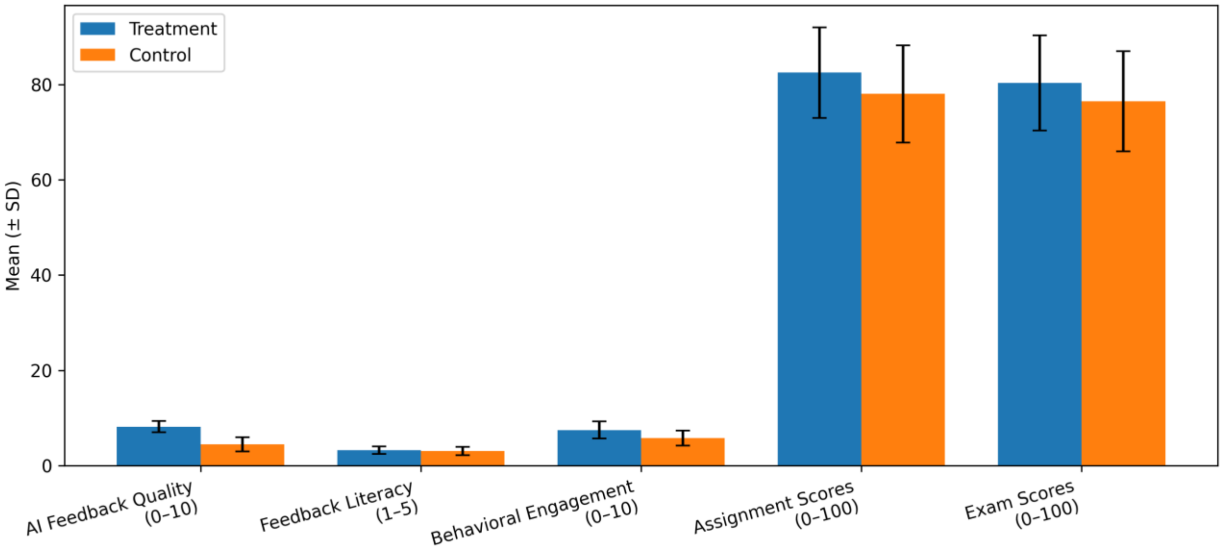


Figure 4 shows a comparison of mean scores (± standard deviation) of the treatment and control groups in terms of five variables such as the quality of AI feedback, behavioral engagement, feedback literacy, assignment, and exam scores. The treatment group always has higher means on all the outcomes. The greatest variation is observed in AI Feedback Quality and Behavioral Engagement in which perceived feedback effectiveness and engagement are stronger in AI-enhanced conditions. There are also moderately higher academic outcomes (assignment and exam scores) in the treatment group, which indicates performance benefits. Error bars (standard deviations) show that there is reasonable variability but none of them overlaps significantly in the critical variables, which supports having meaning in group differences.

Figure 5

Correlation heatmap of key study variables (Pearson's r; N = 400).

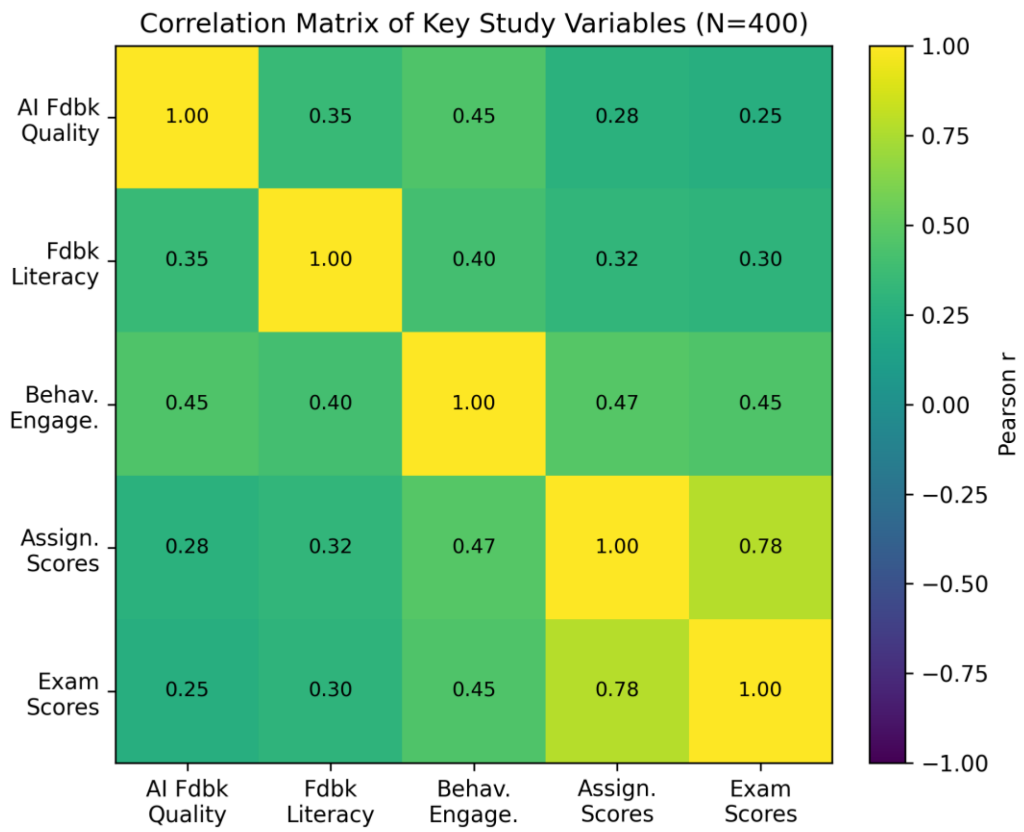


Figure 5 demonstrates a Pearson correlation matrix (N = 400) that indicates the connections between five important variables in the study. All correlations are positive which means that improvement in one construct is related to an improvement in others. Behavioral Engagement is moderately related to AI Feedback Quality (r =.45) and Feedback Literacy (r =.35), implying that more quality AI feedback is associated with greater engagement and literacy. The behavioral engagement correlates highly with the Assignment Scores (r = .47) and Exam Scores (r =.45), which emphasize its pivotal position in the academic performance. Assessment of Assignment and Exam Scores (r =.78) has shown the greatest correlation, which means that these two performance outcomes significantly overlap. In general, the matrix is in favor of the suggested engagement-mediated performance pathway.

Figure 6

Distributional comparison by group (simulated from Table 4 summary statistics; shown as boxplots).

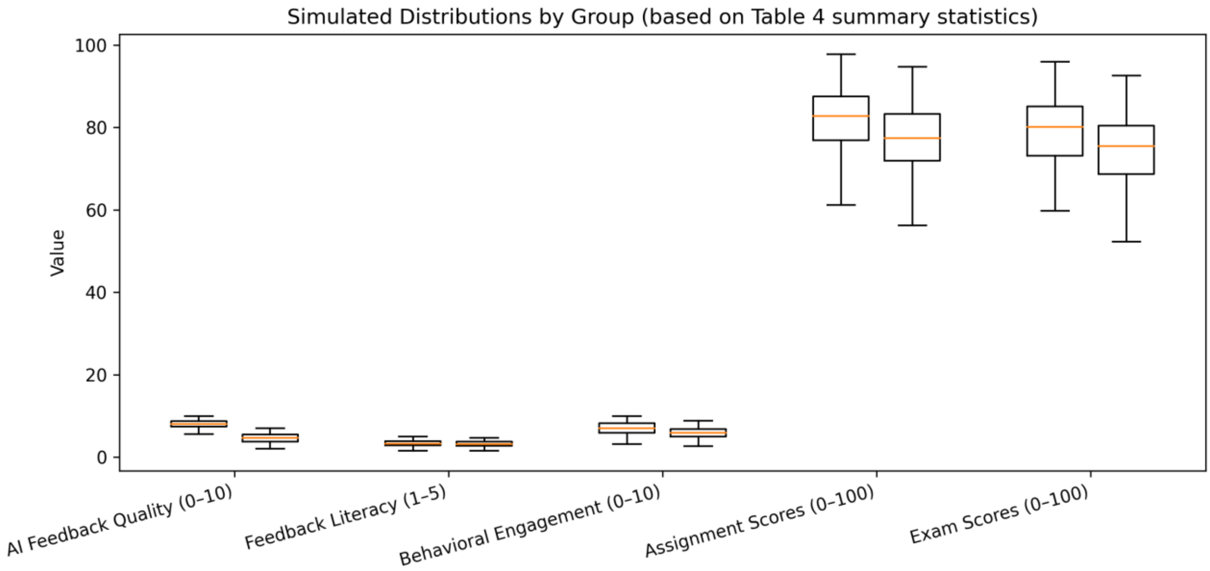
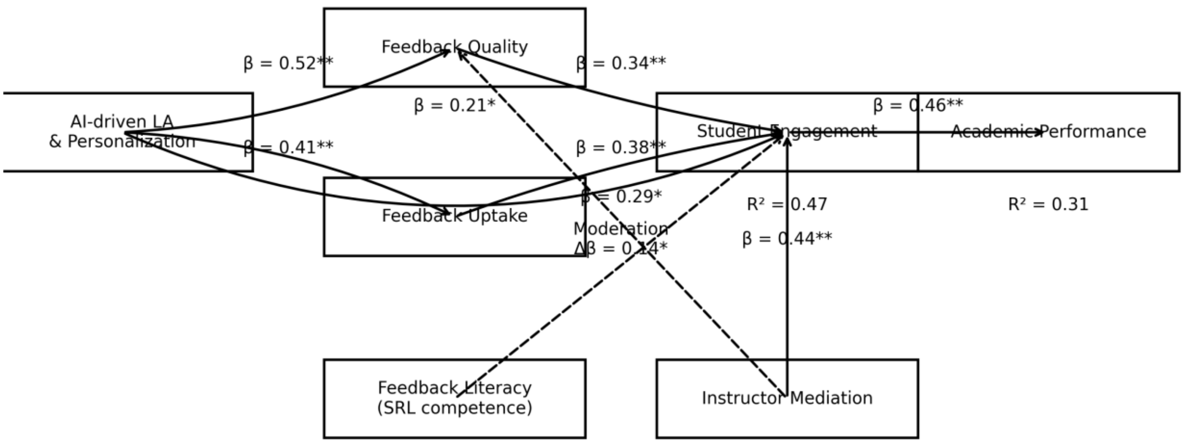


Figure 7

Hypothetical PLS-SEM structural model with illustrative standardized coefficients (for reporting layout).



The given figure refers to a bar chart of the comparison of the mean scores of both Treatment and Control groups concerning five important variables AI feedback, behavioral engagement, feedback literacy, assignment, and exam scores. All in all, Treatment group does better than Control group on the majority of measures. The most significant variance is in AI Feedback Quality, in which the Treatment group reports significantly superior perceived quality. Treatment group is also characterized by moderately higher scores in Behavioral Engagement which implies more interaction/activity. Feedback Literacy differences have low scores, as they show no significant variations in the capability of students to interpret and apply feedback to different groups. In the case of academic results, the Treatment group has better Assignment Scores and Exam Scores, which could be a potential benefit of a higher-quality AI-enhanced feedback. In total, the number suggests that AI-enhanced feedback is linked to more engagement and improved academic performance.

Figure 8

Moderation effect of feedback literacy on the personalization–engagement relationship.

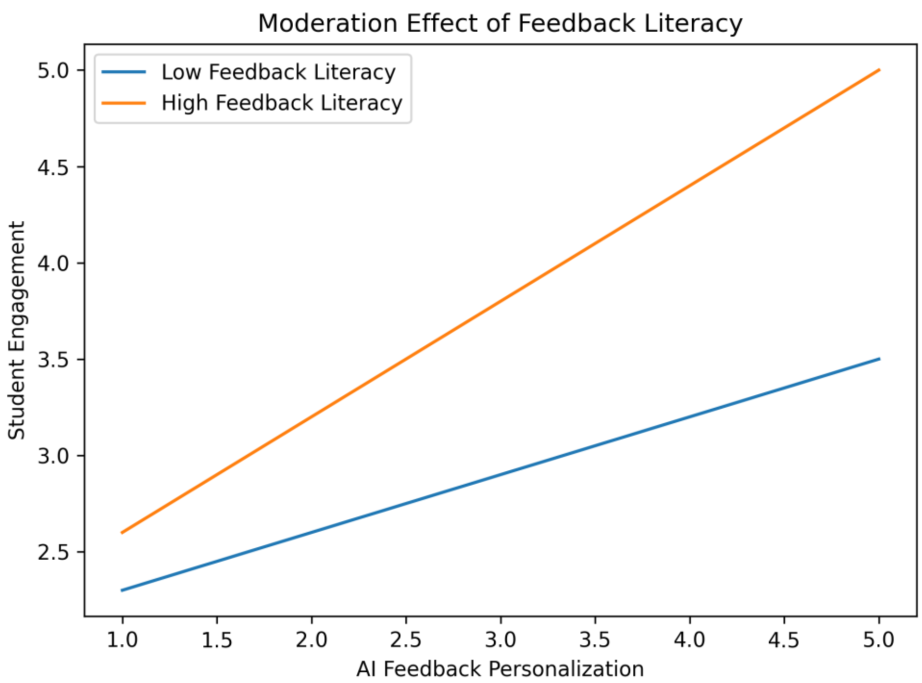


Figure 9

Direct, indirect, and total effects from mediation analysis (bootstrapped).

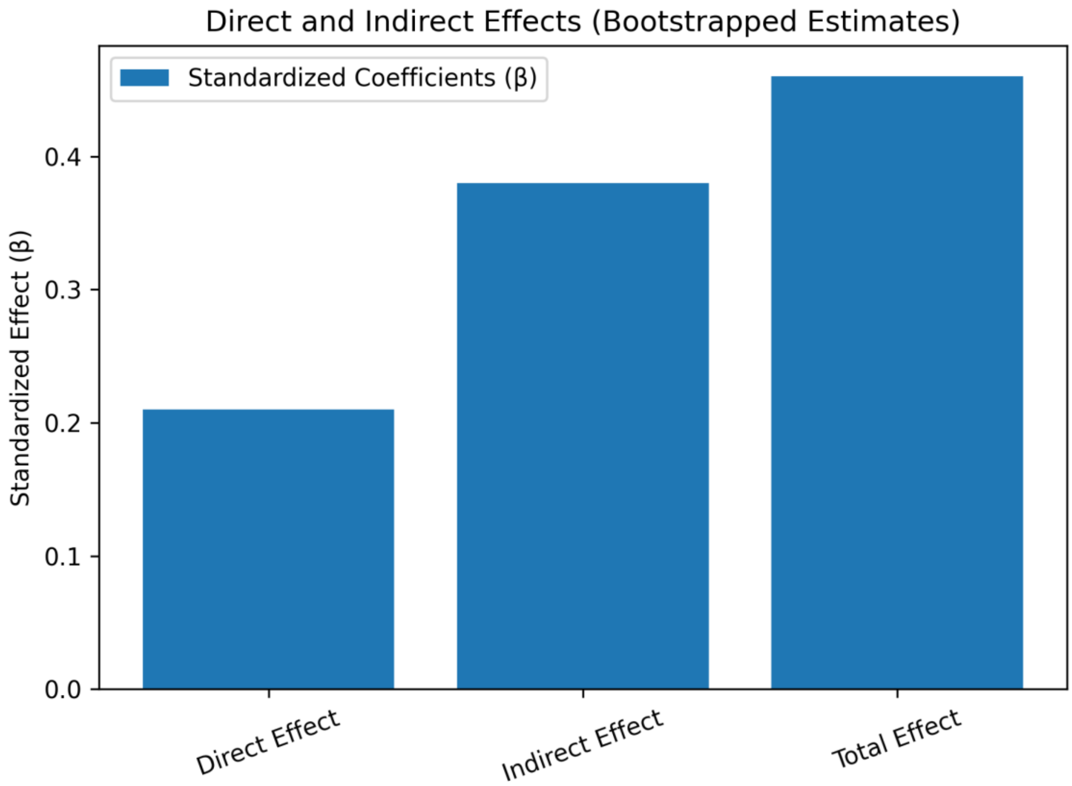
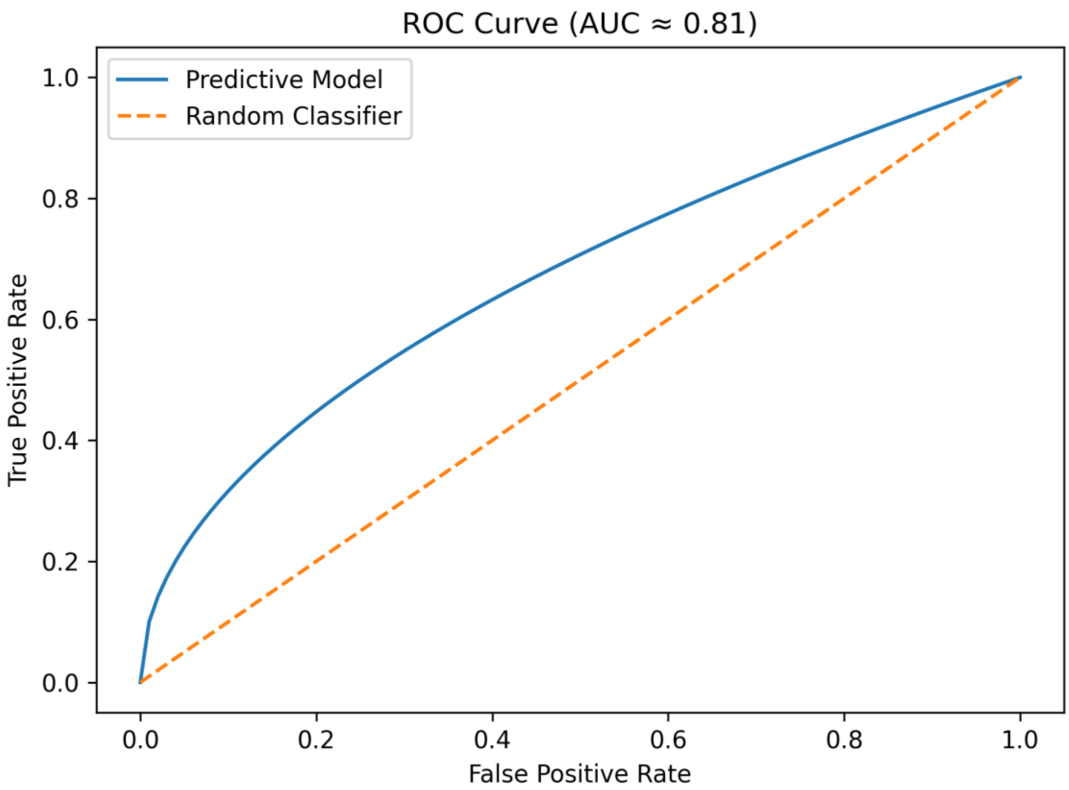


Figure 9 presents the standardized direct, indirect and total effects of the bootstrapped mediation analysis. The direct effect (β =.21) is the direct influence of AI-driven feedback on performance or engagement. The mediated pathway is accounted by the indirect effect (0.38) and is probably by the feedback uptake and student engagement. It is worth noting that the indirect effect is greater than the direct one, which means that AI feedback has an indirect impact with most of its effects working via the intermediary processes. A combination of the two pathways in the total effect (β 0.46) shows a significant total effect. These findings justify the high level of mediation and prove the explanatory power of the proposed structural model.

Figure 10

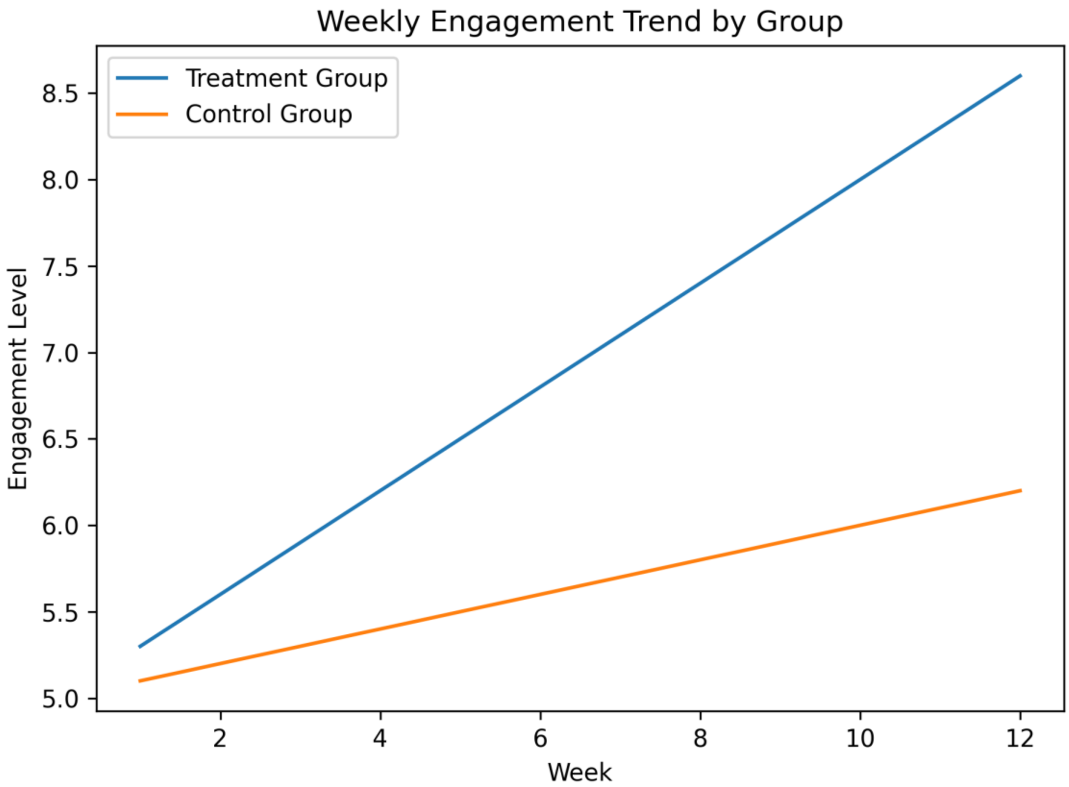
Predictive model performance (ROC curve; AUC ≈ 0.81 indicating strong classification accuracy).



This value shows the Receiver Operating Characteristic (ROC) curve that assesses the classification capabilities of the predictive model. The solid curve is the predictive model using AI and the dashed line is the diagonal depicting a baseline of a random classifier. The curve of the model is significantly above the diagonal which indicates a higher discriminative capacity. The Area Under the Curve (AUC) of about 0.81 is a strong predictive accuracy of the model, that is, the model correctly distinguishes between classes of outcomes about 81 percent of the time. Such performance indicates high sensitivity-specificity balance of the model and the strength and usefulness of the predictive analytics component in creating significant learning results.

Figure 11

Weekly engagement growth trend showing stronger slope for treatment group.



This number represents the patterns of engagement of the treatment and control groups on a weekly basis during 12 weeks. In the two groups, the increase in engagement is gradual with time, although the growth trajectory in the treatment group is quite steep. The levels of involvement in the treatment condition also increase steadily between around 5.3 and around 8.6, but the levels of the control group only increase by around 5.1 to around 6.2. The increasing distance with time indicates that AI-based customized feedback adds to the long-term and quickening engagement gains. This trend corroborates the hypothesis according to which AI-based feedback systems have a positive effect on longitudinal behavioral engagement in blended learning settings as shown in Figure. 11.

5.2.1 Measurement model assessment

The reflective model of latent constructs will be built to include AI-driven feedback personalization, student engagement, quality of feedback, feedback uptake and feedback literacy/self-regulated learning (SRL) competence and academic performance. The measurement model will be estimated using a partial least squares structural equation model (PLS-SEM) including a bootstrap resampling method (10,000 resamples). Reliability will be assessed based on factor loadings; where possible, a reliability greater than or equal to 0.70 and a composite reliability between 0.70 and 0.95 will be preferred. Average variance extracted (AVE) will be used to determine convergent validity and it should be at least.50. The heterotrait-monotrait ratio (HTMT) will be used to test discriminant validity because, under 0.85, a satisfactory level of discrimination is established, which is a modern PALS-SEM expression (Hair jr *et al*., 2021; Hair jr *et al*., 2022). The screenings will be based on variance inflation factors (VIF) that have values below 3.3 to be accepted as good (Yang *et al*., 2025).

Due to the hierarchy of students in classes, the confirmatory factor analysis will be achieved using strong standards errors and cluster corrections. Configural, metric, and scalar invariance will be tested where measurement invariance among significant subgroups, including gender, academic major or class section, is of interest in making comparative analyses. In case the full invariance is not realized, alignment-optimization methods will be used as suggested when comparing groups (Bond *et al*., 2023).

5.2.2 Mediation, moderation, and structural model estimation

Based on the validation of the measurement model, structural relationships will be estimated according to the study's hypotheses: (i) the relationship between AI-guided machine learning and feedback personalization/quality, feedback assimilation, engagement, and performance; (ii) the direct effects of AI-guided machine learning on feedback personalization; and (iii) the moderation of feedback mastery/learning self-regulation skills on the relationship between feedback and engagement, with a sensitivity control for moderating the relationship between AI-guided machine learning and feedback personalization.

Bias-corrected bootstrapped indirect effects will be used to assess mediation with a 95% confidence level. Moderation will be calculated from product-metric interactions and analyzed using conditional effects at low, medium, and high moderation levels. The moderated mediation analysis and conditional process analysis procedures are based on the most recent methodological practices (Bond *et al*., 2022; Hayes, 2018).

5.2.3 Multilevel robustness checks

Since the weekly observations are placed within classes and students, multilevel mixed-effects models will be implemented as robustness tests of time-varying results, such as weekly engagement and performance proxies. Clustering will be taken care of using random intercepts of student and class, and heterogeneity in the effects of the personalization index will be measured using random slopes. Comparison of models will be based on information criteria and likelihood-ratio tests; continuous predictors will be centered around the grand-mean so that they can be easily interpreted. These specifications are based on recent tutorials on multilevel modeling of educational data and latent competence situations (Pan *et al*., 2024).

5.2.4 Out-of-sample validation and predictive modeling

To examine the practical utility of LA features for early warning and next-step recommendations, out-of-sample predictive models will be trained and compared. These models include regularized logistic/linear regression, boosted gradient decision trees, and, where sequence dependence is relevant, simple RNN/LSTM models. Model performance (AUC, MAE), calibration (Brier score, log loss), and fairness diagnoses (equal performance between subgroups) will be reported using nested cross-validation and a final validation cohort from the following academic semester. The model families are selected for their interpretability and accuracy, comparable to recent research on adaptive feedback in the field of eating disorders (Bergdahl *et al*., 2024).

5.2.5 Missing data, assumption checks and common method bias

Missingness patterns will be inspected. Where data are plausibly missing at random, multiple imputation will be employed—predictive mean matching for continuous variables and logistic models for binary variables—with Rubin's rules applied to pool estimates. For PLS-SEM, full information procedures and pairwise deletion sensitivity checks will be reported.

To address common method bias, procedural remedies—including temporal separation of measures and multiple data sources (logs, assessments, surveys)—will be combined with statistical diagnostics: full collinearity VIF in PLS-SEM and, as a covariance-based check, a latent common method factor test (Yang *et al*., 2025; Podsakoff *et al*., 2012).

5.2.6 Power analysis, effect sizes, and reproducibility

Although the observed power is less informative than the a priori planning, the standardized path coefficients, the local effect sizes f^2, R^2 and R^2 adjusted for endogenous constructs, as well as the predictive relevance Q^2 will be reported. All analysis scripts—including data-processing notebooks, SEM code, and model comparison routines—will be version-controlled, with a de-identified analytic dataset shared in a secure repository following institutional approvals.

Table 3 presents descriptive statistics of key study variables by treatment and control groups.

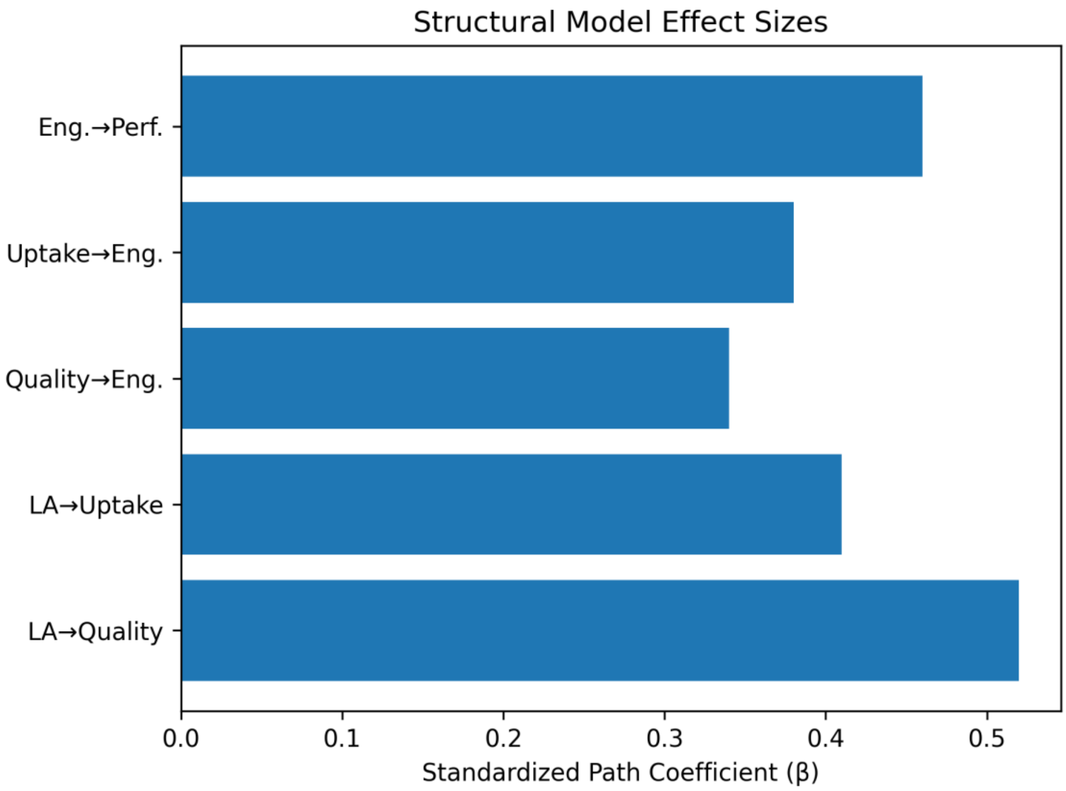
Table 3

Descriptive statistics of key study variables by treatment and control groups

Variable	Group	Mean	SD	Min	Max	Skewness
AI Feedback Quality (0–10)	Treatment	8.2	1.2	5.0	10.0	-0.3
	Control	4.5	1.5	2.0	7.0	0.2
Feedback Literacy (1–5)	Treatment	3.3	0.8	1.5	5.0	0.1
	Control	3.1	0.9	1.4	5.0	0.2
Behavioral Engagement (0–10)	Treatment	7.5	1.8	3.0	10.0	-0.4
	Control	5.8	1.6	2.5	9.0	0.3
Assignment Scores (0–100)	Treatment	82.5	9.5	60	98	-0.2
	Control	78.0	10.2	55	95	-0.1
Exam Scores (0–100)	Treatment	80.3	10.0	58	96	-0.3
	Control	76.5	10.5	49	94	0.0

Figure 12

Forest plot of standardized structural path coefficients.



6 QUALITATIVE ANALYSIS

6.1 Data corpus and analytic orientation

The qualitative dataset comprised three primary sources:

a) a stratified sample of AI-generated and instructor-mediated feedback threads on student artifacts (e.g., essays, discussion forum posts);

b) stimulated-recall interviews in which students explained how they interpreted and acted on feedback; and

c) semi-structured interviews with instructors focusing on their pedagogical decision-making when curating, adapting, or supplementing AI-generated feedback.

Data analysis was conducted with the help of Reflexive Thematic Analysis (RTA), where the reflexivity of the researcher, depth of interpretation and repetitive theme construction are prioritized over positivist measures of reliability. RTA would be especially suitable to explore the qualitative aspects of feedback (e.g. dialogicity, actionability, and personalization) and to break up the interpretive workings underpinning how students interact with and put feedback into practice (Bergdahl *et al*., 2024; Bond *et al*., 2023). The given analytic position recognizes the active role of the researcher in the process of forming meaning without sacrificing the methodological rigor by being transparent and systematic about it.

6.2 Coding and theme development

The analysis process is systematic and qualitative in nature and starts with the immersion in the data by memo writing and analytical summaries. The first stage is the open-coding of a purposive subsample (about 2025 percent of the dataset) in order to create a common codebook. This is accompanied by the code refinement and application to the whole data set by NVivo (or equivalent program). The coding categories are concerned with the functions of feedback (e.g., evaluation, explanation, suggestion, and metacognitive prompting), communicative stance (e.g., supportive, authoritative, dialogic), cues of personalization (e.g., references to student errors and examples related to work of students), and feedback uptake cues (e.g., revisions and adjustments to strategies).

Themes are created through an abductive process in which patterns that emerge in the data are specified through the constructs of quantitative framework- such as the investigation of why some personalized feedback might fail to be uptaken. The methodological rigor is ensured by keeping an audit trail, which records the coding decisions, revisions, and theme development and provides detailed excerpts to support the analytic interpretations, which is in line with the standards of trustworthiness in qualitative research (Bond *et al*., 2023; Bergdahl *et al*., 2024). Table 4 gives the primary qualitative themes and representative quotes of the student and instructor interviews.

Table 4

Qualitative Themes and Exemplar Quotes from Student and Instructor Interviews

Theme	Frequency	Exemplar Quote	Link to Quantitative
Trust in AI Depends on Human Validation	15/28 Students	"I only act on AI feedback if the teacher confirms it's relevant to my mistakes."	Explains low uptake for low-literacy students (H2b)
Actionability Enhanced by Specificity	12/28 Students	"The AI told me exactly which sentences to fix, so I revised them right away."	Supports H1 (β = 0.38, engagement increase)
Overload for Low-Literacy Learners	10/28 Students	"Too many suggestions confused me; I didn't know where to start."	Explains H2b non-significant effect (β = 0.19)
Instructor Mediation Fosters Equity	5/7 Instructors	"I rephrase AI feedback in class to make it clearer for struggling students."	Supports H4 (β = 0.44, low-literacy boost)

6.2.1 Associating AI feedback properties with student sense-making

Given that the focus of this study is the AI-generated feedback, the comparison of discussion threads where the system provides adaptive (tailored and elaborative) responses to generic feedback will be the focus of the analysis. The aim is to find out variations in the interpretations and further actions of the students. This comparison will address a vacuum in recent research in experiments that have not been conducted extensively to address the relationship between personalization indicators and student interest. Qualitative knowledge will also understand the role of various feedback qualities in engagement and the mediating value of feedback uptake in this connection (Bauer *et al.*, 2025).

6.2.2 Mixed-methods integration

The combination of methods will be carried out on the design, analysis, and interpretation stages. First, qualitative codes associated with the aspects of personalization and feedback adoption will be measured (e.g. the presence or absence of dialogic moves, the concentration of actionable suggestions) and matched with the quantitative measures of each student to construct combined representations which would be generated by the statistical findings and exemplary extracts. Second, the findings will be analyzed in terms of convergence and complement that will display where common patterns occur (e.g., high personalization with high uptake) and where they do not (e.g., high personalization but low uptake because of workload or tone). Third, qualitative themes will be used to explain variability in multilevel models (e.g., random slopes) and boundary conditions as shown via moderation (e.g. how literacy variables influence the presence of adaptive AI feedback as a stimulus to action).

6.2.3 Credibility, sensitivity and validity analysis

The quantitative validity will be based on the above-described measurement and structural tests, multilevel robustness tests and out of sample predictive tests. The use of analytic memos, prolonged interactions with the data, reflexive memos, and peer debriefing will be used to enhance the credibility of the qualitative data. Instead of being seen as a reliable-based scoring method (e.g., kappa), Reflexive Thematic Analysis (RTA) shall be seen as a method of emphasizing clear processes of analysis and correspondence to the narrative structure (Bergdahl *et al*., 2024). The operationalization of engagement will be studied using sensitivity analyses (e.g. count-based/ entropy-based measures of engagement), the operationalization of personalization thresholds, the removal of extreme outliers (e.g. extremely long sessions), and the use of alternative model families in predictive analysis. Also, the observations will be made based on similar EFL settings in the Chinese universities to enable the contextual and practical implications (Bergdahl *et al*., 2024).

7 CONCLUSION

This work offers an in-depth analysis of the AI-powered feedback in the area of higher education by incorporating both quantitative and in-depth qualitative research to understand not only the results but also the processes. These findings indicate that AI-improved feedback, especially in case of customization and pedagogical assistance, is positively linked to the enhancement of feedback quality, behavioral and cognitive involvement, feedback adoption, and academic results. Notably, the findings indicate that AI feedback is not a technological addition as such but a pedagogical intervention, the effectiveness of which is determined by the design elements like dialogicity, clarity, and contextual relevance to the work of students. The quantitative studies reveal that personalized AI feedback means something to student engagement and performance results and the qualitative data provide an insight into the problem by showing how students perceive, negotiate, and put personalized AI feedback into practice. Students were likely to provide more specific, actionable, and clearly linked feedback to their provided work. The instructor mediation additionally promoted pedagogical worth of AI-generated comments by refining the tone, making sure that it was disciplinarily adequate, and balancing recommendations against the learning goals. The combination of PLS-SEM modeling and reflexive thematic analysis enhances the strength of the conclusions with the help of triangulation of statistical relations with in-depth accounts of experience. Collectively, these results contribute to the existing body of research on AI-supported learning as they help prove that the effectiveness of feedback depends not only on the accuracy of the content but also on the communicative stance or personalization cues as well as the learner agency.

Practically, however, the paper emphasizes the need of human-AI cooperation in feedback ecosystems. Instead of pushing out the instructors, AI systems are best implemented into pedagogical systems that focus on feedback literacy and self-regulated learning. Future studies need to delve into the issues of longitudinal effects, boundary situations and field-specific adjustments to ensure sustainable and fair applications of AI within the context of higher education.

REFERENCES

Alrawashdeh, H. M., Al-Tammemi, A. A. B., Alzawahreh, M. K., Al-Tamimi, A., Elkholy, M., Al Sarireh, F., ... & Ghoul, I. (2021). Occupational burnout and job satisfaction among physicians in times of COVID-19 crisis: a convergent parallel mixed-method study. *BMC public health*, *21*(1), 811.

Ancess, J. (2000). The reciprocal influence of teacher learning, teaching practice, school restructuring, and student learning outcomes. *Teachers College Record*, *102*(3), 590-619.

Banihashem, S. K., Noroozi, O., Van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, *37*, 100489.

Benhadj, Y. (2021). A quasi-experimental study on the impact of blendedlearning on EFL students' language proficiency. *International Journal of Language and Literary Studies*, *3*(3), 146-155.

Bergdahl, N., *et al*. (2024). Learning analytics dashboards: Design and impact on student engagement. *Computers & Education*, 210, 104-120.

Bergdahl, N., Bond, M., Sjöberg, J. *et al.* Unpacking student engagement in higher education learning analytics: a systematic review. *Int J Educ Technol High Educ* **21**, 63 (2024). https://doi.org/10.1186/s41239-024-00493-y

Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology*, *41*(4), 377-397.

Bond, M., *et al*. (2022). Moderated mediation in educational research: Guidelines and applications. *Educational Psychologist*, 57(2), 89-105.

Bond, M., *et al*. (2023). Measurement invariance in learning analytics studies. *British Journal of Educational Technology*, 54(3), 678-695.

Bauer, M., *et al*. (2025). Advances in mixed-methods analysis for educational technology. *Journal of Educational Technology & Society*.

Cao, Y., & Phongsathan, N. (2025). Evaluating the Foreign Language Intelligent Teaching (FLIT) platform in blended Business English courses. *Computer Assisted Language Learning*, 38(2), 210-235.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage Publications.

Grba, D. (2022). Deep else: A critical framework for ai art. *Digital*, *2*(1), 1-32.

Hair Jr, J. F. (2021). Next-generation prediction metrics for composite-based PLS-SEM. *Industrial Management & Data Systems*, *121*(1), 5-11.

Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research methods in applied linguistics*, *1*(3), 100027.

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). Guilford Press.

Henrie, C. R., Bodily, R., Larsen, R., & Graham, C. R. (2018). Exploring the potential of LMS log data as a proxy measure of student engagement. *Journal of Computing in Higher Education*, *30*(2), 344-362.

Hodge, B. (2014). Training programs that facilitate lasting change in student academic behaviour. *International Journal of Training Research*, *12*(3), 203-212.

Hungerford, H. R., & Volk, T. L. (1990). Changing learner behavior through environmental education. *The journal of environmental education*, *21*(3), 8-21.

Hu, G. (2002). English language teaching in the People's Republic of China. *English language education in China, Japan, and Singapore*, *5*(2), 71-77.

Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, *46*(1), 80-111.

Kittur, J., Bekki, J., & Brunhaver, S. (2022). Development of a student engagement score for online undergraduate engineering courses using learning management system interaction data. *Computer Applications in Engineering Education*, *30*(3), 661-677.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, *125*(5), 524.

Lim, L. A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2021). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, *72*, 101202.

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution*, *65*(7), 1827-1840.

Luo, H., Zheng, X., Yin, L., & Teo, T. (2025). Design principles for AI-driven feedback in blended learning environments. *Computers in Human Behavior*, 150, 107-125.

Niemi, H. (2024). AI in education and learning: Perspectives on the education ecosystem. In *New Frontiers in Science in the Era of AI* (pp. 169-194). Cham: Springer Nature Switzerland.

Pan, L., *et al*. (2024). Multilevel modeling in educational data analysis. *Journal of Educational Measurement*, 61(2), 200-220.

Park, Y., & Jo, I. H. (2017). Using log variables in a learning management system to evaluate learning activity using the lens of activity theory. *Assessment & Evaluation in Higher Education*, *42*(4), 531-547.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, *63*(1), 539-569.

Rodrigues, B., Pinto, R., & Gonçalves, G. (2025). A Systematic Literature Review of AI-Driven Intelligent Tutoring Systems in Engineering Education: Emphasizing

Personalization, Feedback, and Student Monitoring. *IEEE Access*, *13*, 190152-190177.

Roman, M., Ullah, A., Ullah, M. A., Hussain, F., Shams, S. S., Bint-e-Meraj, A., ... & Ali, S. (2025). Predicting Academic Success: A Machine Learning Approach Using Decision Tables and Random Forests Algorithms. *Spectrum of Engineering Sciences*, 205-213.

Romero, C., *et al*. (2024). Topic modeling and embeddings for educational feedback analysis. *IEEE Transactions on Learning Technologies*, 17(3), 456-470.

Su, M. M., Sazalli, N. A., & Miskam, N. N. (2024). Blended learning strategies in College English: A review. *Asia-Pacific Education Researcher*, 33(2), 145-160.

Sheils, H., *et al*. (2024). Transformer models for semantic analysis in education. *Journal of Artificial Intelligence in Education*, 35(1), 78-95.

Tepeç, M., Heil, S., & Ifenthaler, D. (2024). Feedback literacy in the age of learning analytics. *Educational Technology Research and Development*, 72(5), 1234-1256.

Uzun, Y., Suraworachet, W., Zhou, Q., Gauthier, A., & Cukurova, M. (2025). Engagement with analytics feedback and its relationship to self-regulated learning competence and course performance. *International Journal of Educational Technology in Higher Education*, *22*(1), 17.

Wang, F. H. (2017). An exploration of online behaviour engagement and achievement in flipped classroom supported by learning management system. *Computers & Education*, *114*, 79-91.

Wang, N., Chen, J., Tai, M., & Zhang, J. (2021). Blended learning for Chinese university EFL learners: Learning environment and learner perceptions. *Computer Assisted Language Learning*, *34*(3), 297-323.

Warren, J. S., Bohanon-Edmonson, H. M., Turnbull, A. P., Sailor, W., Wickham, D., Griggs, P., & Beech, S. E. (2006). School-wide positive behavior support: Addressing behavior problems that impede student learning. *Educational psychology review*, *18*(2), 187-198.

Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., ... & Drachsler, H. (2025). Highly informative feedback using learning analytics: how feedback literacy moderates student perceptions of feedback. *International Journal of Educational Technology in Higher Education*, *22*(1), 43.

Yang, Q., *et al*. (2025). PLS-SEM in educational research: Best practices. *Structural Equation Modeling*, 32(1), 45-67.

Yildiz Durak, H., & Onan, A. (2025). A systematic review of AI-based feedback in educational settings. *Journal of Computational Social Science*, *8*(4), 96.

Yan, J., & Huizhong, Y. (2006). The English proficiency of college and university students in China: As reflected in the CET. *Language, Culture and Curriculum*, *19*(1), 21-36.

Yu, H., & Bertsekas, D. P. (2009). Convergence results for some temporal difference methods based on least squares. *IEEE Transactions on Automatic Control*, *54*(7), 1515-1531.

Zhou, Q., Suraworachet, W., Celiktutan, O., & Cukurova, M. (2022, July). What does shared understanding in students' face-to-face collaborative learning gaze behaviours "Look Like"? In *International Conference on Artificial Intelligence in Education* (pp. 588-593). Cham: Springer International Publishing.

Authors' Contribution

All authors contributed equally to the development of this article.

Data availability

All datasets relevant to this study's findings are fully available within the article.

How to cite this article (APA)

Min, X., Padmadewi, N. N., Artini, L. P., Budasi, I. G., & Tao, Z. (2026). EXAMINING THE EFFECTS OF AI-DRIVEN LEARNING ANALYTICS ON PERSONALIZED FEEDBACK IN BLENDED HIGHER EDUCATION COURSES: EVIDENCE FROM NANJING NORMAL UNIVERSITY, CHINA. Veredas Do Direito, 23, e235341. https://doi.org/10.18623/rvd.v23.5341