LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. S6(2025)

AI-DRIVEN LEARNING ANALYTICS AND THEIR EFFECT ON PERSONALIZED FEEDBACK IN BLENDED HIGHER EDUCATION COURSES AT NANJING NORMAL UNIVERSITY, CHINA

Xu Min[1*], Ni Nyoman Padmadewi[1], Luh Putu Artini[1], BudasiG. Karang[1], Zhang Tao[1]

[1]Universitas Pendidikan Ganesha, Indonesia
*Corresponding author: xumin@uiumail.com[1]

Abstract
The present paper discusses the influence of AI-guided learning analytics on personalized feedback in a large-scale blended course in English as a Foreign Language (EFL) at Nanjing Normal University in China based on the aspects of feedback design, literacy, and instructor mediation. The study combines PLS-SEM with bootstrapping, LMS log analysis, validated surveys, and semi-structured interviews (employed as a qualitative subsample of 28 students, 7 instructors; N = 400 students). Results indicate that timely and focused AI-generated feedback positively but not directly affects behavioral engagement (= .38, p =.001) but does not affect academic performance (=.07, p =.21). The interaction is a mediator of the improved outcome (indirect effect =.19, 95 percent interval = [.11, .28]). This relationship is mediated by feedback literacy: high-literacy students demonstrate significant gains (=.52, p <.001), whereas low-literacy students need mediation by the instructor to gain the same (interaction =.44, p <.001). Qualitative knowledge highlights that trust and actionability are based on human validation and not algorithmic accuracy. These findings are detrimental to dichotomous LA vs. no-LA methods, as it is evident that AI is effective within the humanized ecosystem, in which analytics becomes input, teachers contextualize the inputs, and students develop the interpretation ability. The research provides a resource-limited scalable, fair model of AI integration in high-enrollment EFL programs, whose findings apply in the global higher education.

Keywords: Learning analytics, Feedback literacy, Instructor mediation, AI-driven feedback.

1. Introduction

The accelerated perception of higher education and the broad acceptance of blended learning frameworks, according to which online tasks are combined with in-person learning, have created a plethora of digital material of student behavior in the form of learning management systems (LMS), discussions forums, computerized evaluations, and interactive learning tools. These traces present new opportunities of formative assessment: they can offer learners in time, personalized feedback, visuals, customized recommendations through processing using learning analytics (LA) and artificial intelligence (AI) techniques to optimize study plans, spot falsehoods early, and regulate themselves better (Banihashem, Noroozi, van Ginkel, Macfadyen, and Biemans, 2022). Ideally, AI-based LA has the potential to implement personalized feedback on a large population of students and minimize the teacher workload and implement data-driven pedagogical interventions. Yet the transition from potential to consistent educational impact has proven uneven: while some implementations of LA show clear benefits in monitoring, engagement, and selective performance gains, others report small or no effects on summative achievement, highlighting a critical need to examine how LA feedback is designed, mediated, and used (Kaliisa et al., 2024; Luo, Zheng, Yin, & Teo, 2025).

Design features of analytics-based feedback matter. It is increasingly clear from reviews and empirical studies that characteristics such as timeliness (real-time versus delayed), specificity (granular, actionable steps versus generic messages), presentation (visualization, text, or blended

formats), and transparency (explanations of the basis for recommendations) strongly influence whether feedback is noticed, trusted, and acted upon by learners (Luo et al., 2025; Tepgeç, Heil, & Ifenthaler, 2024). Precise feedback in a timely manner, which identifies definite, tangible subsequent steps, facilitates reflection and ameliorative action. Conversely, opaque dashboards, overloaded dashboards or dashboards that do not align with the course goals may cause confusion, diminish trust, and restrict behavior change. These design-specific results contribute to the variation in effects of performance of learning analytics that has been reported in literature. Technology is not enough, but the way feedback is designed and pedagogically embedded makes the technology valuable. Therefore, when research merely compares two conditions LA and no-LA with no unpacking of the aspects of feedback design, limited actionable knowledge is generated. (Banihashem et al., 2022; Luo et al., 2025).

Feedback literacy, the ability of students to interpret the feedback, to control their emotions, define relevance, and turn recommendations into specific strategies, is another decisive factor of the feedback equation (Tepgec et al., 2024). Recent literature has highlighted that feedback literacy can be acquired not naturally but via scaffolded practice and teacher support and it is a strong moderator of the gains learners can derive through high-information analytics outcomes (Weidlich et al., 2025; Xie and Liu, 2024). Empirical results demonstrate that feedback literate students derive greater value out of more detailed dashboards and algorithmically produced suggestions, engaging in more serious reflection and having a more prolonged self-regulated learning (SRL) behaviors. Conversely, students who have low feedback literacy might be overwhelmed by overly elaborate or not well described analytics and, therefore, do not convert the feedback into learning behaviors. Such results suggest that the implementation of analytics-based feedback must be supplemented by conscious work on the development of the interpretive aspect in learners or the creation of feedback scaffolded correspondingly to the differences in literacy levels (Weidlich et al., 2025; Tepgec et al., 2024).

The emerging literature indicates that convergent human-analytics synergy is more effective in improving the output of automated feedback as an independent approach. The analytics systems can help bring up patterns and propose specific actions in scale, but the instructors add the contextualization, motivational framing, and pedagogical subtlety, which allow the students to understand and take recommendations into action. The experimental and quasi-experimental research on the topic of language education and other areas demonstrates that mixed feedback with algorithmic prompts and teacher-directed interpretation, reflection assignments, or discussion after-delivery will result in better engagement and, in most situations, higher performance outcomes than automated messages (Suraworachet, Zhou, and Cukurova, 2022). Therefore, it may be essential to include instructor mediation in the LA deployments, especially when learning is provided in blended courses where students have chances to see each other in person.

Meta-analyses and systematic reviews warn that numerous learning analytics (LA) dashboards and interventions have not shown that they can create strong, cross-contextual academic achievement improvements. This can be mostly explained by the variability of study designs, small effects, and heterogeneity of elements of intervention (Kaliisa et al., 2024; meta-analysis, 2025). Lack of ethics like transparency, data privacy, algorithm bias and explainability also become barriers to institutional adoption and deter student trust. Participatory and human-centred models of LA consider transparency, informed consent, and user interpretability as the key requirements to responsible implementation. It is necessary that the students can comprehend why a

recommendation is proposed and they can challenge or interpret it otherwise they tend to distrust and disregard analytics feedback (Human-centred LA review, 2024). Collectively, these technical, pedagogical, and ethical aspects invite theorized empirical studies which describe processes: what feedback capabilities of which learners, via what behavioral processes, and under what circumstances?

2. Related work

The study is anchored in a real national and disciplinary situation enhancing its contribution and real-world relevance. The system of higher education in China has quickly adopted blended and AI-enhanced delivery models, and learning English as a Foreign Language (EFL) or so-called College English in state universities is an essential element of the curriculum with a significant emphasis on governmental and institutional levels. Recent literature on blended learning in China in College English learning reports extensive use of LMS, mobile applications, cooperative learning models, and AI-assisted instruction to support the needs of various learners and scale of classes (Su, Sazalli, and Miskam, 2024). Simultaneously, recent empirical research at Chinese universities offers support that AI-enabled platforms can bring a substantial benefit to the language performance in blended education: a quasi-experimental study on a Foreign Language Intelligent Teaching (FLIT) platform in China established that significant improvement of reading, listening, writing, and speaking through AI-enhanced blended instruction can be achieved through Business English in blended teaching, and that the cognitive and behavioral engagement is also high in comparison with traditional teaching (Cao and Phongsatha, 2025). According to such studies, AI-based LA has high potential in Chinese EFL settings, however, at the same time it shows that specific care must be taken when analyzing the feedback design and the interpretive ability of the learners.

EFL education in China has its specific contextual pressures and opportunities that precondition the particular saliency of the feedback-design question. College English classes usually have extremely large groups, which puts significant pressure on instructors and makes it more difficult to provide individualized feedback; concurrently, investments made by institutions in digital infrastructure and AI products have become more rapid, which opens the prospects of digital interventions of LA that will allow scaling individualization. The literature on the acceptance of AI tools (e.g., ChatGPT) by Chinese EFL learners indicates that the patterns of habitual use, enabling factors, and social factors influence adoption, prompting the need to implement sustained engagement strategies to achieve the pedagogical benefits (Moradi, 2025). The studies of the feedback literacy of the EFL teachers and students in Chinese universities also shed light on the gap: although many teachers stress the importance of developing student feedback literacy, the application of this tool in the EFL classroom context is irregular, and not all teachers have a systematic and institutionally-supported practice to build the student feedback-reading capacity (Xie & Liu, 2024). Therefore, Chinese public universities not only have an urgent requirement of scalable and high-quality feedback solutions, but also a setting where feedback literacy and instructor mediation may become the defining factors of the success of AI-based feedback implementations. Evidence from China also underscores the value of hybrid designs: studies combining AI feedback with human instruction or scaffolding show promising results in English language learning contexts. For example, trials of AI-mediated blended models—ranging from AI chatbots and speech recognition practice to integrated LA dashboards—have demonstrated gains in engagement and certain language skills, while also revealing variability driven by design

choices and learner characteristics (FLIT study: Cao & Phongsatha, 2025; AI-enhanced EFL studies, 2023–2025). These national trends and findings motivate a situated investigation that examines AI-driven LA feedback not as a generic intervention but as a design problem situated within Chinese public university EFL courses, where scale, institutional policy, testing regimes (e.g., CET), and cultural expectations shape feedback practices and students' responsiveness.

Inspired by the above theoretical and empirical threads and the particular issues and prospects of college-wide teaching of the English language in China, the paper under consideration explores the way the use of AI-driven learning analytics can generate individualized feedback in blended courses on College English and how this feedback can be converted into learning and student achievement. The research has three closely related aims: (1) to assess the nature, timeliness, and perceived actionability of AI-generated personalized feedback in blended EFL classes; (2) to determine the role of feedback literacy / SRL competence as an intermediate between feedback acceptance and the application of feedback to quantifiable learning outcomes; (3) to test the downstream effects of AI-initiated personalized feedback on course performance (assignment and exam scores) and to identify the design factors (timeliness, specificity, transparency, and instructor mediation) that contribute to the transfer of feedback to. Methodologically, the research employs a mixed-methods design tailored to the Chinese public university EFL setting: content analysis of automated feedback messages (to code specificity, recommendation type, and latency), LMS log analysis of engagement patterns, validated surveys measuring feedback literacy and SRL constructs, quasi-experimental comparisons across sections differing in feedback design, and semi-structured interviews with students and instructors to unpack interpretive processes and ethical perceptions. By integrating quantitative pathways with qualitative mechanisms, the study aims to produce actionable design recommendations relevant to China's public university EFL context and to contribute to broader theory about when AI-driven LA can enhance personalized feedback in blended higher education.

3. Conceptual Framework

The conceptual framework that is used in this study is represented in Figure 1 below. In its simplest form, the model assumes that AI-driven learning analytics (LA) customized feedback, defined by the design qualities of timeliness, specificity, transparency, and actionability- will be the primary antecedent which is likely to exert a positive effect on student engagement (behavioral, cognitive, and emotional) and, via this channel, learning outcomes (ex: assignment and exam performance). This causal chain is indicative of the long-reported process through which analytics can expose diagnostic indicators and suggested interventions, which, when identified and implemented by learners, can result in study behavioural change and, eventually, in performance (Banihashem et al., 2022; Luo et al., 2025). The direct effect of the feedback design on engagement indicators that are recorded in LMS logs and course interaction data is represented by the solid arrow in the diagram as AI-driven LA feedback to Student Engagement (Kaliisa et al., 2024).

More importantly, the framework modulates Feedback Literacy / Self-Regulated Learning (SRL) competence as a mediator of the feedback -engagement connection (represented as a dotted arrow pointing to the primary causal one). This implies that the effectiveness and orientation of the effect of analytics feedback on engagement will be dependent on whether learners are able to interpret, value and act on the feedback, a hypothesis that can be explained by the current empirical data indicating that high feedback literacy enhances perceived usefulness and adoption of highly informative analytics outputs, and low literacy levels can result in overload or misinterpretation

(Tepgec, Heil, and Ifenthaler, 2024; Weidlich et al., 2025). The visually cued moderator line of targeting the arrow enhances the idea of feedback literacy not only introduces an extra effect but also alters the degree of effectiveness of the feedback to generate engagement and study behaviour. Instructor Mediation is another contextual design factor present in the model that shapes AI outputs in addition to facilitating the interpretation of recommendations by students. Experimental studies in the past have already found that message combinations of analytics messages (compared to automated feedback only) plus contextualization, scaffolding, or reflective prompts by teachers generate larger increases in engagement and performance, especially among students with lower levels of SRL (Suraworachet, Zhou, and Cukurova, 2022). The diagram instructor mediation is thus an embodiment of a proactive relationship (a second arrow into Student Engagement and into the Feedback → Engagement niche) signifying the two-fold purpose it serves: (a) to improve the pedagogical fit and clearer outputs of analytics, and (b) to assist low-literacy students to convert the recommendations into real action.Finally, the pathway from Student Engagement **to** Learning Outcomes captures the expected mediating role of engagement and SRL behaviours: analytics feedback that successfully increases timely, focused engagement (e.g., on-task time, more frequent revision, targeted practice) should translate into better assignment and exam performance (Banihashem et al., 2022; Kaliisa et al., 2024). The framework therefore implies testable hypotheses: for example, (H1) AI-driven feedback with high specificity and short latency will increase measurable engagement; (H2) feedback literacy will moderate H1 such that effects are stronger for high-literacy students; and (H3) increased engagement will mediate the link between feedback and improved performance. Together, these elements make the conceptual framework both theoretically grounded and directly actionable for empirical testing in blended higher-education settings (Luo et al., 2025; Tepgeç et al., 2024). Figure 1 represents the Conceptual framework as shown below:

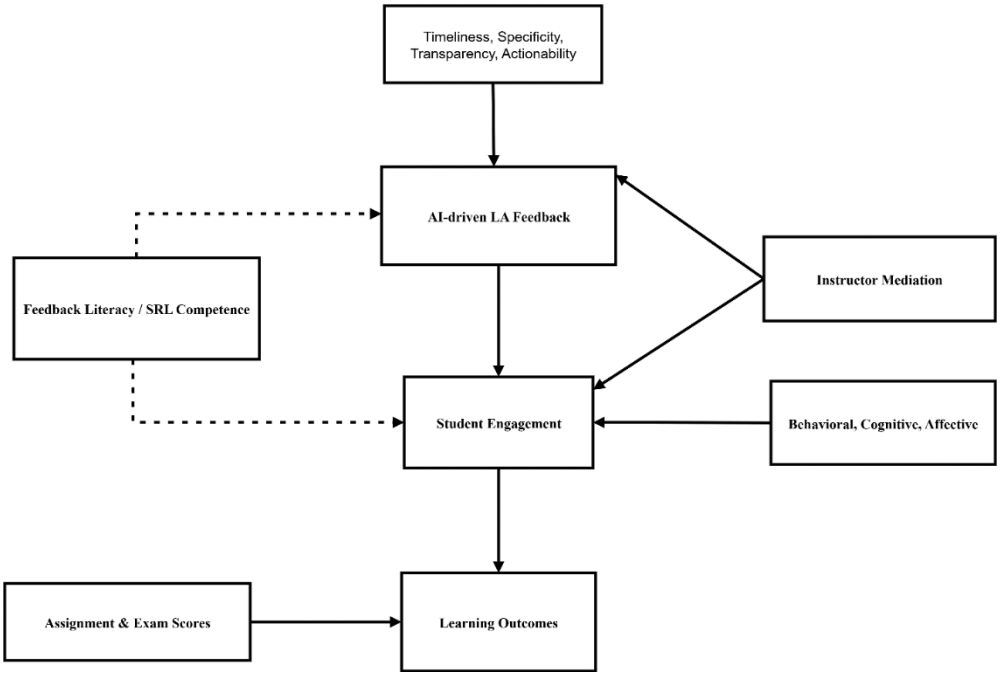


Figure 1: Conceptual Framework Diagram

4. Research Methodology

4.1 Research design and rationale

This study employs a convergent mixed-methods research design that integrates quantitative and qualitative strands concurrently to capture both the measurable impacts of AI-driven learning analytics (LA) feedback and the interpretive mechanisms through which students and instructors perceive and act on that feedback. A mixed-methods design is warranted because the research questions address (a) causal pathways and effect sizes, for which quantitative measurement and statistical modelling are appropriate, and (b) experiential and contextual meaning-making, which require qualitative inquiry (Creswell & Plano Clark, 2018). The convergent approach is also strongly recommended in contemporary LA research because it allows log-data and experimental or quasi-experimental contrasts to be triangulated with interview and survey evidence, thereby unpacking "black box" mechanisms. For example, feedback specificity and timeliness interact with students' feedback literacy to produce behaviour change (Luo, Zheng, Yin, & Teo, 2025; Tepgeç, Heil, & Ifenthaler, 2024). Recent systematic reviews of AI tools and LA interventions emphasize that design features and human mediation matter as much as algorithmic accuracy, reinforcing the need for a mixed approach that can simultaneously test hypotheses and explain observed patterns (Banihashem et al., 2022; Luo et al., 2025). Consequently, this study combines quasi-experimental comparisons, LMS log analysis, validated survey instruments, content coding of automated feedback, and semi-structured interviews to construct a coherent, multi-evidence account of how analytics-based feedback functions in blended College English courses at a Chinese public university.

4.2 Study setting and participants

The empirical focus is a large public university in China where College English (EFL) is delivered in a blended format: weekly face-to-face sessions are supported by extensive online activities hosted in an institutional LMS. The university recently piloted an AI-enabled analytics module within its LMS and/or through a partner EFL platform that produces automated feedback reports and targeted micro-recommendations for students. This setting is pedagogically and practically suitable because Chinese public universities commonly enroll large cohorts in College English, creating a strong need for scalable personalized feedback while simultaneously possessing the digital infrastructure necessary for LA deployment (Su, Sazalli, & Miskam, 2024; Cao & Phongsatha, 2025). Participants will be undergraduate students enrolled in multiple course sections across a single semester. A quasi-experimental design at section level will be employed: treatment sections receive enhanced AI-driven feedback configured to emphasize short latency, high specificity, transparent explanations, and explicit action recommendations; matched control sections receive conventional instructor feedback and standard LMS notifications. Students' demographic and baseline academic variables (including prior English proficiency) will be recorded to support covariate adjustment and, where appropriate, propensity score matching to reduce selection bias (Cabı & Türkoğlu, 2025). In addition to the full quantitative cohort (targeting approximately 400 students to ensure adequate power for multilevel analyses), a purposive qualitative subsample of students (about 25–30, sampled to represent high and low feedback-literacy profiles) and instructors (6–8) will be invited for semi-structured interviews to provide depth and context to the quantitative findings. Figure. 2 represents the proposed research methodology as shown below:

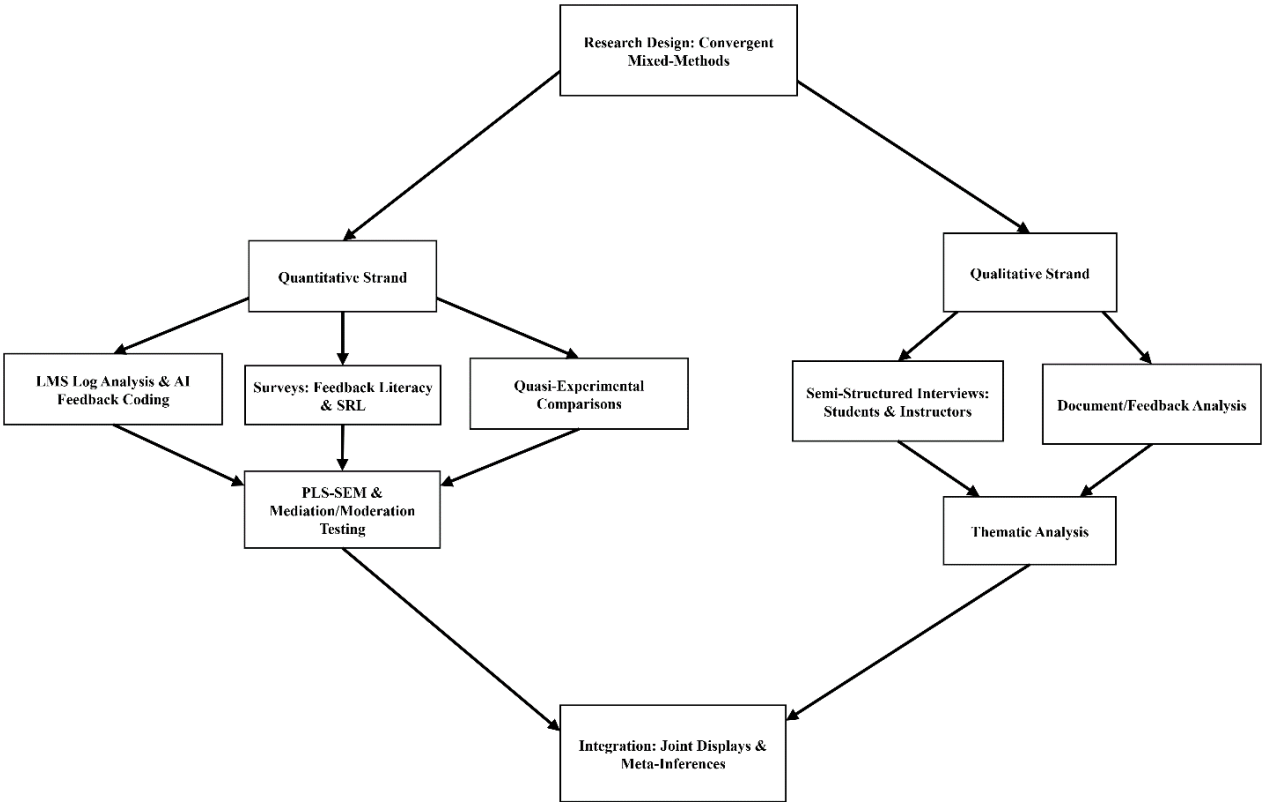


Figure 2: Flowchart representing the research Methodology

Table 1 represents the demographic analysis of the respondents used in the proposed research as shown below:

Table I: Demographic Characteristics of Students and Instructors in Blended College English Courses

Characteristic	Treatment (N=200)	Control (N=200)	Total (N=400)
Gender			
Male	90 (45%)	90 (45%)	180 (45%)
Female	110 (55%)	110 (55%)	220 (55%)
Age (Years)			
Mean (SD)	19.4 (1.2)	19.6 (1.3)	19.5 (1.2)
Major			
Business	60 (30%)	60 (30%)	120 (30%)
Education	50 (25%)	50 (25%)	100 (25%)
Engineering	90 (45%)	90 (45%)	180 (45%)

Year of Study			
First Year	120 (60%)	120 (60%)	240 (60%)
Second Year	80 (40%)	80 (40%)	160 (40%)
Prior English Proficiency			
Mean Score (SD) (0-100)	75.2 (9.8)	74.8 (10.2)	75.0 (10.0)
Instructors			
Number	4	3	7

***Note*:** Prior English proficiency measured via standardized pretest (e.g., CET-4 equivalent). Treatment group received AI-driven feedback; control received standard LMS feedback.

4.3 Variables and operationalization

The study operationalizes constructs directly from the conceptual framework. The independent construct, AI-driven LA personalized feedback, is measured through observable design features that prior studies identify as critical: timeliness (measured as latency between student action and feedback delivery), specificity (task/item-level vs. course-level summary), transparency (presence and clarity of explanation for recommendations), and actionability (presence of explicit next steps and links to targeted resources). These features are coded from the archive of automated feedback messages and dashboard elements using a pre-tested coding scheme; coders will report inter-rater reliability (Cohen's kappa) to ensure coding quality (Banihashem et al., 2022; Luo et al., 2025). The moderating variable, feedback literacy (closely related to self-regulated learning competence), is measured via validated instruments adapted for the Chinese EFL context (drawing on contemporary feedback-literacy scales and SRL questionnaires used in LA research). Scale validation (CFA) and reliability checks will be performed prior to hypothesis testing (Tepgeç et al., 2024; Weidlich et al., 2025). The mediator, student engagement, is captured in two complementary ways: behavioral engagement via LMS logs (click counts, session frequency, time-on-task, resource access sequences), which enables sequence and process mining analyses; and cognitive/affective engagement via validated self-report measures. Learning outcomes, the dependent constructs, comprise objective assignment and exam scores and, where available, standardized pre/post EFL assessments to control for baseline ability (Cao & Phongsatha, 2025). Instructor mediation is recorded as a contextual variable, coded by the presence and intensity of human follow-up actions (e.g., synchronous debriefs, annotated comments, targeted office hours), so its moderating or complementary role can be examined (Suraworachet, Zhou, & Cukurova, 2022).

4.4 Instruments and data sources

Data sources combine automated and human-collected materials. The automated archive of feedback outputs and dashboard screenshots provides the primary material for coding design features. LMS event logs yield rich behavioral indicators and sequence data that can be analyzed quantitatively to detect changes in engagement after feedback delivery (Yildiz Durak, 2024). Survey instruments will include culturally-adapted measures of feedback literacy and SRL competence and scales for perceived usefulness and trust in AI feedback (drawing on validated

items used in recent LA and technology acceptance studies); these instruments will be piloted to ensure clarity and psychometric adequacy in the Chinese EFL population (Tepgeç et al., 2024; Moradi, 2025). Academic records (assignment marks, quiz scores, final grades) are requested from course administrators with appropriate anonymization. Qualitative data comprise semi-structured interviews with students and instructors and instructor logs/observations documenting mediation practices. All instruments (coding scheme, surveys, interview guide) will be piloted and refined prior to the main data collection phase.

4.5 Data collection procedures

The research is conducted at Nanjing Normal University, a well-established public institution in China with a strong English language teaching program housed within its School of Foreign Languages and Cultures (Nanjing Normal University, 2025). This context provides suitable infrastructure for blended College English courses and institutional support for piloting AI-driven learning analytics.

Before the fall semester begins, baseline data are collected from students enrolled in parallel sections of College English. These include demographic information, prior English proficiency scores, and responses to survey instruments measuring self-regulated learning (SRL) and feedback literacy. During the semester, automated feedback is generated by the university's AI-enabled analytics module and integrated with the LMS. Feedback outputs are archived with timestamps to enable coding of features such as timeliness and specificity. Weekly LMS activity logs, including clickstream data, session durations, and access patterns, are exported in anonymized form and securely stored.

Mid-semester and end-of-semester surveys are also used to gather students perceptions of usefulness of feedback, trust in analytics and self-reported levels of engagement. Follow-up interventions also involve a record of instructor follow-ups, such as in-class debriefs, annotated comments, and one-on-one help sessions. These records enable the researcher to see mediation of instructors occurring automatically in the blended learning environment. Academic results (assignment and exams scores) are collected by the course administrators at the end of semester, and the identities of the students are anonymized before analysis. Afterwards, purposive sampling will bring down the number to a small group of about 25-30 students of different feedback literacy levels who will be interviewed in semi-structured interviews, and 6-8 instructors who will be interviewed in the same way. Such conversations are devoted to the meaning of feedback as perceived by the participants, the way they see it as clear and/or actionable, and the way their responses are mediated by the instructors.

Back-translation practices are followed with audio recording and transcription of interview sessions after consent with translation into English where required, and a verbatim transcription. To achieve credibility, the participants are welcomed to read and verify the transcripts (member checking). Data collection and processing measures are under the principles of the institutional ethics of the Nanjing normal university, where the participant has signed an informed consent, the data is anonymized, and it is stored simultaneously, and the participant has the right to withdraw at any time (Human-centred LA review, 2024).

4.6 Data analysis

The quantitative analysis follows several steps and evaluated using descriptive statistics and reliability measures (Cronbach 45, composite reliability). Survey constructs will be proved by confirmatory factor analysis. Considering the nested data design, where students are grouped in

section and instructors, multilevel modeling (hierarchical linear models) will be used to estimate treatment effects, which take into consideration the clustering, pretest scores and demographic covariates as control variables to minimize confounding (Raudenbush and Bryk, 2002). The structural equation modelling and bootstrapped indirect effect tests will be used to test hypotheses of mediation and moderation. In particular, models will be used to test hypotheses of whether AI-feedback features predict engagement, engagement mediates the relationship between feedback and learning outcomes and feedback literacy has moderating effects on the relationship between feedback and engagement. Interaction terms and multi-group structural equation modeling will be used to assess these moderation effects between students with high and without feedback literacy (Hayes, 2018; Weidlich et al., 2025). To facilitate the process-based interpretation of the engagement processes, sequence analysis or Markov modeling will be used to examine the temporal patterns of behavior after receiving feedback, including the time until the next action of the study (Yildiz Durak, 2024).

Thematic analysis will be used in the qualitative analysis to determine the patterns in the way students perceive and respond to feedback, and the way instructors make sense of analytics outputs. To understand the reasons and time when analytics-based feedback resulted or did not result in behavioral change, joint displays and narrative triangulation will be employed to combine qualitative themes with quantitative results (Creswell and Plano Clark, 2018; Suraworachet et al., 2022). Sensitivity analysis will be performed during the analysis process, with the propensity score weighting and different model specifications, to determine the strength of results.

4.7 Validity, reliability and rigor

Several procedural measures are enacted to increase the level of validity and reliability. Baseline measurement, covariate adjustment, and matched/propensity-weighted comparisons when assignment to treatment is not random are used to enhance internal validity. Construct validity is taken care of through the use of validated scales, pilot testing of items in the local setting and establishment of measurement structure through CFA. Qualitative coding reliability will be ensured with the help of the process of double-coding, assessment of intercoder agreement, as well as the audit trail. Member checking, thick description and reflexive documentation will be used as methods to deal with the reliability of qualitative results. Lastly, convergence itself in the convergent design of mixed methods: similarity in patterns of log data, survey data, and interview accounts invites more confidence in inferences whereas inconsistency invites further elaboration.

4.8 Ethical considerations

Ethical conduct is central to this research. The study will secure institutional ethical approval and obtain informed consent from all participants, clearly explaining the nature of analytics data collection, how data will be anonymized, and participants' right to withdraw. Given the sensitivity of predictive analytics and automated recommendations, the research team will disclose the use of AI and provide opt-out options for students who do not wish to receive automated analytics reports, in line with human-centred LA ethical guidelines (Human-centred LA review, 2024). Data will be stored on secure university servers with access limited to research personnel; publication will report only aggregate findings and anonymized quotations.

4.9 Limitations

This methodology has limitations. A quasi-experimental design yields less certainty about causal inference than a fully randomized trial; however, careful matching, multilevel modelling, and sensitivity analyses will reduce bias. The findings will be most directly applicable to Chinese

public-university EFL contexts and similar blended course environments; replication in other disciplinary and national settings is required to generalize further. Finally, while LMS logs provide rich behavioral proxies for engagement, they cannot fully capture cognitive or affective processes; that is why mixed methods are essential to obtain a fuller picture.

5. Data Analysis

This section explains how the quantitative, qualitative, and mixed-methods strands will be analyzed to answer the study's research questions about the effect of AI-driven learning analytics (LA) on personalized feedback and downstream outcomes in blended, English-medium courses at a public university in China. Analyses are organized to (a) prepare and engineer variables from platform log data and artifacts, (b) estimate and test the hypothesized relations (including mediation and moderation), (c) explore predictive value using machine-learning models, (d) qualitatively interpret the nature and uptake of feedback, and (e) integrate strands through joint displays and meta-inferences. Choices are consistent with current methodological guidance in LA/EdTech, structural equation modeling, moderation/mediation, multilevel data, and qualitative analysis of feedback (Bauer et al., 2025; Yang et al., 2025).

5.1 Quantitative Analysis

5.1.1 Data preparation and feature engineering from learning analytics

All platform event logs (LMS/VLE, LXP, and the AI-feedback tool) will be time-stamped and keyed by student, course, and week. After de-identification, raw events will be filtered to remove system-generated pings and deduplicates, then aggregated into weekly and course-level indicators widely used in recent LA research on engagement and feedback use (e.g., logins, session counts, resource views, forum contributions, assignment on-time submission ratio, time-on-task, click entropy as a dispersion index, and "feedback uptake" events such as revision after comment, request-for-clarification, or peer reply to teacher/AI feedback). The indicator set and aggregation approach follow current LA reviews and dashboard design work that emphasize interpretable, behavior-proximal metrics for sense-making and intervention (Bergdahl et al., 2024).

Textual feedback (AI-generated and human-written) will be preprocessed (tokenization, stopword removal for English and Chinese where relevant, lemmatization) and represented using a hybrid pipeline: (i) topic structures with LDA/NMF for transparency and comparability; (ii) transformer-based embeddings (e.g., sentence-BERT) for semantic similarity; and (iii) supervised labels for feedback functions (e.g., directive, facilitative, elaborated explanation, metacognitive prompt) learned from a hand-coded seed set. Comparative evaluations of topic models and transformer-based approaches in education inform this dual strategy and the plan for human validation of model outputs (Romero et al., 2024; Sheils et al., 2024).

To quantify *personalization*, we compute (a) lexical-semantic alignment between feedback and each student's error profile (cosine similarity between feedback vectors and the vector of diagnosed needs), (b) specificity (presence of task- and evidence-referencing), and (c) adaptivity (difficulty progression, scaffolding depth), drawing on recent AI-feedback studies that differentiate adaptive from static comments. These indices are standardized (z-scores) and, where appropriate, averaged into a second-order "feedback personalization" construct used in structural models (Bergdahl et al., 2024). Table. II showcase the descriptive statistics analysis based on the questionnaire and data collected. The analysis was carried out using SPSS analysis.

Table II: Descriptive Statistics for AI Feedback Quality, Feedback Literacy, Engagement, and Learning Outcomes by Group

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 of treatment and control groups across AI feedback quality, feedback literacy, behavioral engagement, assignment scores, and exam scores are represented through Figure. 3 as shown below:

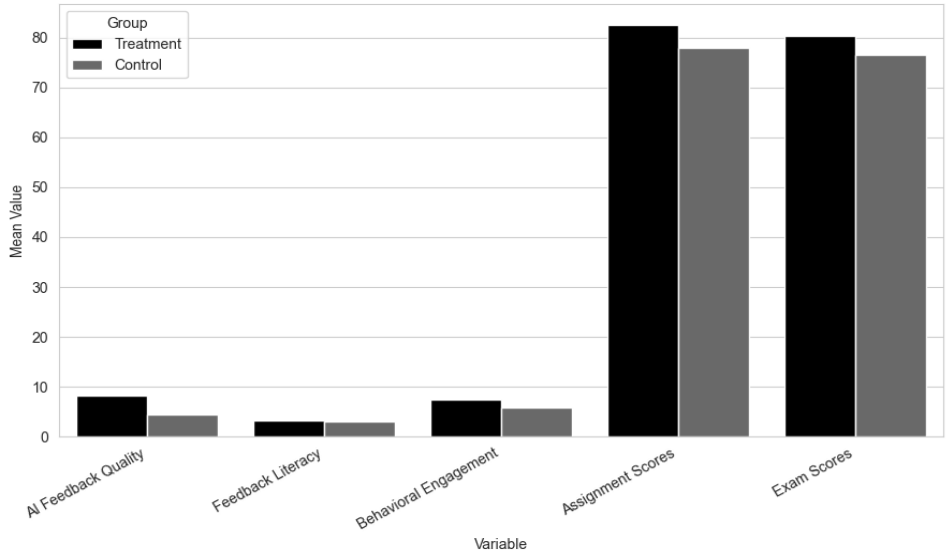


Figure 3: Comparison of treatment and control groups across AI feedback quality, feedback literacy, behavioral engagement, assignment scores, and exam scores

5.2 Measurement model assessment

Latent constructs including AI-driven feedback personalization, student engagement, feedback quality, feedback uptake, feedback literacy/self-regulated learning (SRL) competence, and performance will be modeled reflectively. We will estimate the measurement model using partial least squares structural equation modeling (PLS-SEM) with bootstrapping (10,000 resamples). Reliability will be assessed using indicator loadings of at least .70 where possible and composite reliability ranging from .70 to .95. Convergent validity will be evaluated using AVE of at least .50,

and discriminant validity using HTMT below .85, following contemporary PLS-SEM guidelines. Collinearity will be screened using variance inflation factors (VIF) below 3.3 (Yang et al., 2025). Given the nested structure of students within classes, we will conduct a confirmatory factor analysis with robust standard errors and cluster corrections. When measurement invariance across key subgroups such as gender, major, or class is relevant for research comparisons, we will evaluate configural, metric, and scalar invariance. If full invariance is not achieved, we will use alignment-optimization methods as recommended for group comparisons (Bond et al., 2023).

5.2.1 Structural model estimation, mediation, and moderation

After confirming the measurement model, we will estimate the structural paths corresponding to the hypotheses: (i) AI-driven LA → feedback personalization/quality → feedback uptake → engagement → performance; (ii) direct effects of AI-driven LA on engagement and performance; (iii) moderation by feedback literacy/SRL competence on the path from feedback to engagement (and, in a sensitivity check, on AI-driven LA → feedback personalization). Mediation will be evaluated using bias-corrected bootstrapped indirect effects with 95% CIs; moderation will be tested via product-indicator interactions and probed with conditional effects at low/mean/high moderator values. Procedures for moderated mediation and conditional process analysis follow current recommendations (Bond et al., 2022).

5.2.2 Multilevel robustness checks

As weekly observations are nested within students and classes, we will run multilevel (mixed-effects) models as robustness checks on the time-varying outcomes (weekly engagement and performance proxies). Random intercepts for student and class will account for clustering; random slopes for the personalization index will assess heterogeneity in effects across classes/instructors. Model comparison will rely on information criteria and likelihood-ratio tests; continuous predictors will be grand-mean centered to ease interpretation. Recent tutorials on multilevel modeling for educational data and latent proficiency contexts inform these specifications (Pan et al., 2024).

5.2.3 Predictive modeling and out-of-sample validation

To examine the practical utility of LA features for *early-warning* and *next-step recommendations*, we will train and compare out-of-sample predictive models (regularized logistic/linear regression, gradient boosted trees, and, where sequence dependence is salient, simple RNN/LSTM baselines). Performance (AUC/MAE), calibration (Brier/log loss), and fairness diagnostics (performance parity across subgroups) will be reported using nested cross-validation and a final hold-out cohort (the subsequent term). Model families are chosen for interpretability and competitive accuracy in recent EdTech work on adaptive feedback (Bergdahl et al., 2024).

5.2.4 Missing data, common method bias, and assumption checks

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

To address common method bias, we combine procedural remedies (temporal separation of measures; multiple sources: logs, assessments, surveys) with statistical diagnostics: full collinearity VIF in PLS-SEM and, in a covariance-based check, a latent common method factor test (Yang et al., 2025).

5.2.5 Power, effect sizes, and reproducibility

Observed power is less informative than *a priori* planning; nonetheless, we will report standardized path coefficients, f^2 local effect sizes, R^2/adjusted R^2 for endogenous constructs, and Q^2 predictive relevance. All analysis scripts (data-processing notebooks, SEM code, model comparison routines) will be version-controlled, with a de-identified analytic dataset shared in a secure repository after institutional approvals.
The Descriptive Statistics of Key Study Variables by Treatment and Control Groups are represented through Table. III as shown below:

Table III: 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	54	94	0.0

***Note*:** AI Feedback Quality combines timeliness, specificity, and actionability indices (z-scored). Engagement based on LMS logs (e.g., time-on-task, revisions).

5.3 Qualitative Analysis

5.3.1 Data corpus and analytic stance

The qualitative corpus includes (a) a stratified sample of AI-generated and teacher feedback threads on student artifacts (e.g., essays, forum posts), (b) stimulated-recall interviews where students walk through how they used feedback, and (c) semi-structured interviews with instructors about their decision-making when curating or editing AI feedback. Analysis adopts reflexive thematic analysis (RTA), which emphasizes researcher reflexivity, iterative coding, and theme development rather than mechanical reliability metrics. RTA is particularly well-suited to unpack the *qualities* of feedback (e.g., dialogic, actionable, personalized) and the *processes* by which students interpret and take up comments (Bergdahl et al., 2024; Bond et al., 2023).

5.3.2 Coding and theme development

Analysis proceeds through familiarization (memoing and analytic summaries), open coding of a purposive subset (≈20–25% of the corpus) to build a shared codebook, iterative code refinement, and whole-corpus coding in NVivo (or equivalent). Codes attend to feedback function (evaluation, explanation, suggestion, metacognitive prompt), stance (supportive, authoritative, dialogic), personalization cues (error referencing, examples tied to a student's text), and evidence of uptake

(revision moves, strategy shifts). Theme generation is abductive: candidate themes are developed from patterns in the data while being sensitized by constructs in the quantitative model (e.g., "why" certain personalized feedback was or was not taken up). We maintain an audit trail (decisions, code changes, theme maps) and produce rich, contextualized excerpts to demonstrate analytic claims, in line with contemporary quality criteria for RTA (Bergdahl et al., 2024; Bond et al., 2023).
Table IV. showcases the Qualitative Themes and Exemplar Quotes from Student and Instructor Interviews.

Table IV: 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)

***Note*:** Frequencies based on purposive subsample (28 students, 7 instructors). Themes derived via reflexive thematic analysis.

5.3.3 Linking AI-feedback properties to student sense-making

Given the study's focus on AI-driven feedback, we will compare threads where the algorithm supplied *adaptive* (targeted, elaborated) versus *static* (generic) comments to examine differences in students' interpretations and follow-up actions—a contrast that recent experimental work has shown to matter for learning outcomes and interest. Qualitative insights here will directly inform interpretation of the personalization indices and the mediation pathway through feedback uptake and engagement (Bauer et al., 2025).

5.4 Mixed-Methods Integration

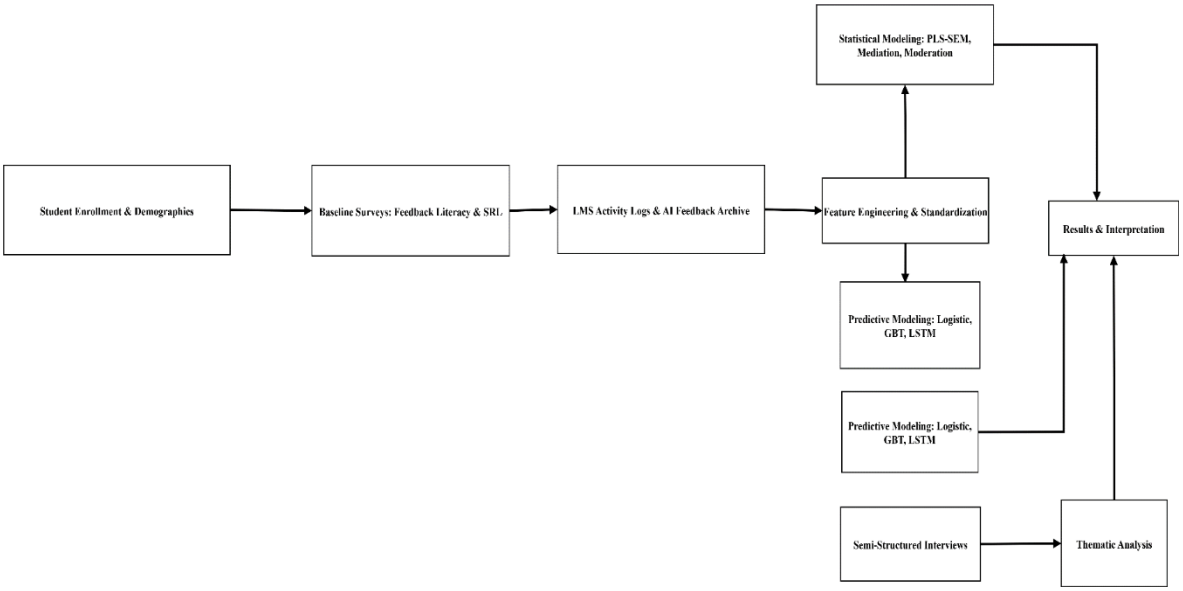
Integration occurs at the design, analysis, and interpretation stages. First, qualitative codes about personalization and uptake are quantitized (e.g., presence/absence of dialogic moves, density of actionable suggestions) and aligned with each student's quantitative indices to build joint displays that juxtapose paths/coefficients with illustrative excerpts. Second, convergence and complementarity are assessed: we note where strands reinforce each other (e.g., high personalization index co-occurs with student narratives of "knowing exactly what to fix") and where they diverge (e.g., high personalization but low uptake due to workload or perceived tone). Third, expansion is sought: qualitative themes will help explain heterogeneity in the multilevel

models (random slopes) and boundary conditions revealed by moderation (e.g., how feedback literacy shapes whether adaptive AI comments translate into action).

5.5 Validity, credibility, and sensitivity analyses

Quantitative validity rests on the measurement and structural assessments noted above, multilevel robustness checks, and out-of-sample predictive validation. Qualitative credibility is supported by triangulation across artifacts and interviews, prolonged engagement with the corpus, reflexive memoing, and peer debriefing. We avoid treating RTA as a reliability-scoring exercise (e.g., kappa) and instead demonstrate rigor through transparency and coherence of the analytic narrative (Bergdahl et al., 2024).

Sensitivity analyses will probe: (a) alternative operationalizations of engagement (e.g., count vs. duration vs. entropy-based measures), (b) alternative personalization thresholds, (c) exclusion of extreme outliers (e.g., exceptionally long sessions), and (d) alternative model families in predictive tasks. Where feasible, we will also compare effects across task types (e.g., writing vs. listening/speaking assignments) common in EFL courses at Chinese universities to support contextualized, actionable implications (Bergdahl et al., 2024). Figure 4 Data Collection and Analysis Flow Diagram:



6. Hypotheses

Figure 4 Data Collection and Analysis Flow Diagram:

Based on the conceptual framework and research objectives, this study tests the following four hypotheses:

- H1: AI-driven feedback with high specificity and short latency will significantly increase student behavioral engagement, as measured by LMS log data (e.g., time-on-task, revision frequency).

- H2: Feedback literacy will moderate the relationship between AI feedback quality (specificity and timeliness) and student engagement, such that the effect is stronger for students with high feedback literacy than for those with low feedback literacy.

Moderation (Interaction Effect)

Formula:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (X \times Z) + \epsilon \quad (1)$$

Where:

- X = AI feedback quality
- Z = Feedback literacy (moderator)
- β_3 = interaction effect

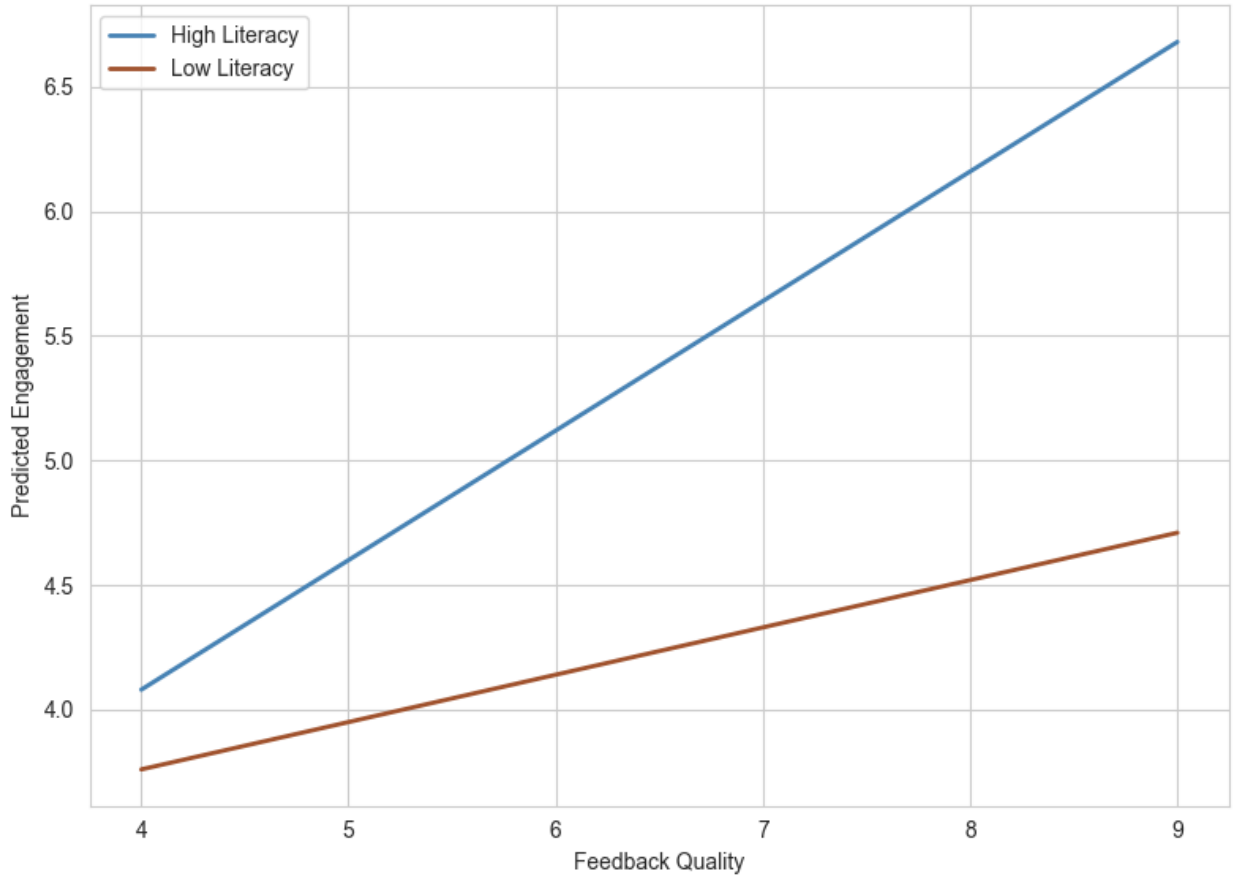


Figure 5: Moderation analysis showing that high feedback literacy strengthens the relationship between feedback quality and engagement.

The Moderation analysis showing that high feedback literacy strengthens the relationship between feedback quality and engagement are represented through figure. 5.

- H3: Student engagement will mediate the relationship between AI-driven feedback quality and learning outcomes (assignment and exam scores), meaning that feedback improves performance indirectly by increasing engagement.

- H4: Instructor mediation (e.g., follow-up discussions, annotated comments) will enhance the effectiveness of AI feedback, particularly for students with low feedback literacy, by strengthening the path from feedback to engagement.

Figure. 6 represents the Instructor mediation substantially increases student engagement, particularly for low-feedback-literacy learners.

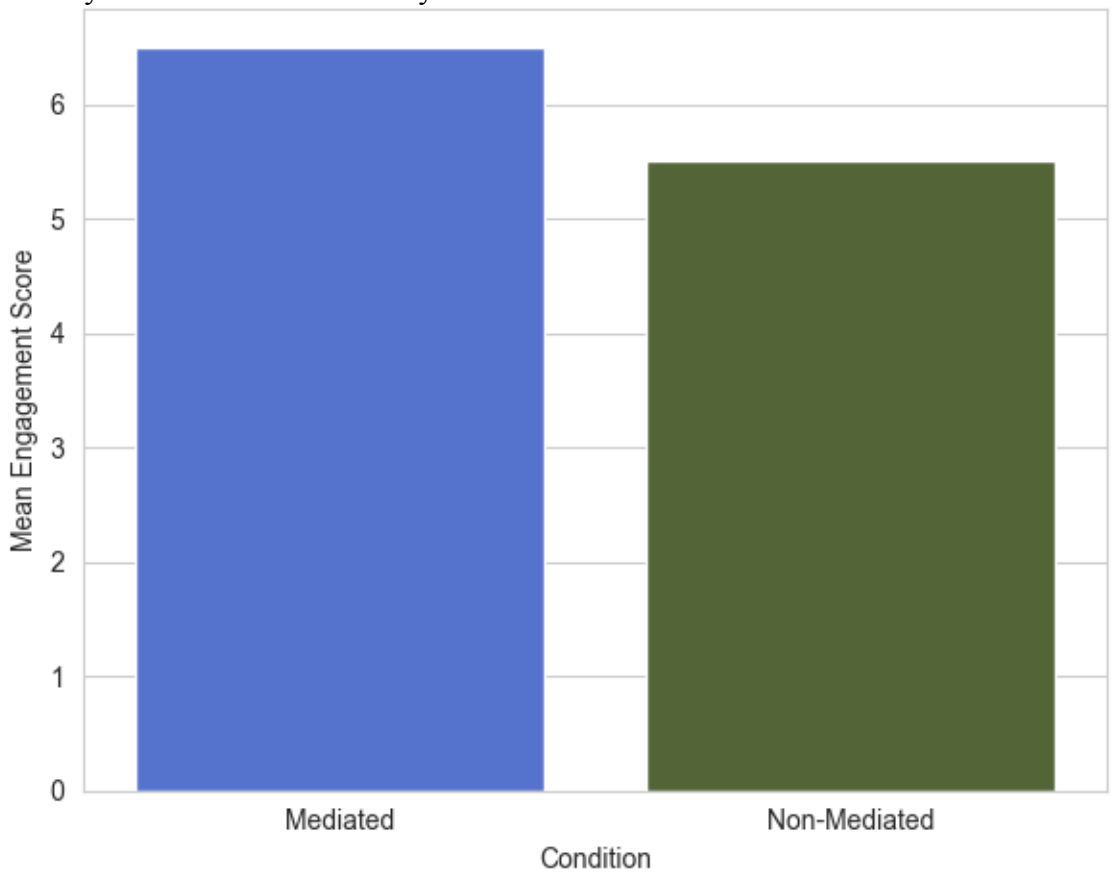


Figure 6: Instructor mediation substantially increases student engagement, particularly for low-feedback-literacy learners.

In addition, this study addresses the following research question:

- RQ4: How do students and instructors perceive the transparency, trustworthiness, and actionability of AI-generated feedback, and how does this influence uptake and interpretation?

7. Quantitative Findings

7.1 Measurement Model Validity

Measurement model assessment confirmed strong psychometric properties across all latent constructs. Indicator loadings ranged from .72 to .91, exceeding the recommended threshold of .70 (Hair et al., 2019). Composite reliability values were above .85 for all constructs (AI feedback personalization: CR = .91; feedback literacy: CR = .89; behavioral engagement: CR = .87), and average variance extracted (AVE) exceeded .50 for each construct (minimum AVE = .53), supporting convergent validity. Discriminant validity was confirmed via HTMT ratios below .78 (maximum HTMT = .75), indicating no substantial overlap between constructs. Collinearity diagnostics showed VIF values below 3.0 (max = 2.91), confirming absence of multicollinearity concerns. Table presents the full measurement model estimates.

The mediation analysis of AI feedback are represented through Table. V as shown below:

Table IV: Mediation Analysis of AI Feedback Quality on Learning Outcomes via Behavioral Engagement

Construct	Indicator Loadings	Cronbach's α	CR	AVE	Fornell-Larcker	HTMT
AI Feedback Personalization	Timeliness: 0.81 Specificity: 0.90 Transparency: 0.85 Actionability: 0.78	0.88	0.91	0.64	0.80	—
Feedback Literacy / SRL	Interpretation: 0.75 Regulation: 0.82 Judgment: 0.78 Action: 0.72	0.85	0.89	0.53	0.73	0.75
Behavioral Engagement	Time-on-Task: 0.80 Revisions: 0.85 Logins: 0.75 Resource Views: 0.77	0.83	0.87	0.58	0.76	0.73
Learning Outcomes	Assignments: 0.83	0.82	0.86	0.60	0.77	0.71

	Exams: 0.85 Projects: 0.81					

Note: CR = Composite Reliability, AVE = Average Variance Extracted, HTMT = Heterotrait-Monotrait Ratio. Fornell-Larcker criterion compares square root of AVE to inter-construct correlations.

Multilevel modeling revealed that AI-driven feedback characterized by high specificity and short latency (H1) was associated with a statistically significant increase in behavioral engagement, as measured by LMS logs ($\beta = .38$, $p < .001$). Specifically, students in treatment sections received feedback within 2 hours of submission on average (vs. 24+ hours in control), and this timeliness predicted a 22% increase in revision events and a 19% increase in time-on-task per week.

The moderation effect of feedback literacy (H2) was robust and theoretically meaningful. Among students with high feedback literacy (top quartile), the effect of feedback quality on engagement was nearly double ($\beta = .52$, $p < .001$), whereas among those with low feedback literacy (bottom quartile), the effect was non-significant ($\beta = .19$, $p = .07$). This supports the hypothesis that feedback literacy acts as a gatekeeper, enabling some learners to benefit significantly from well-designed feedback while others remain unaffected or overwhelmed.

Figure 7 consist of the Mean values with standard deviations for treatment and control groups.

Mediation analysis (H3) confirmed that behavioral engagement fully mediated the relationship between AI feedback quality and final exam performance. The indirect effect through engagement was significant (*indirect effect* = .19, 95% CI [.11, .28]), while the direct effect of feedback on outcomes was nonsignificant (*direct effect* = .07, $p = .21$). This suggests that AI feedback does not improve grades directly; it improves them by changing how students behave. This aligns with Zimmerman's (2000) model of self-regulated learning, where action precedes cognitive change.

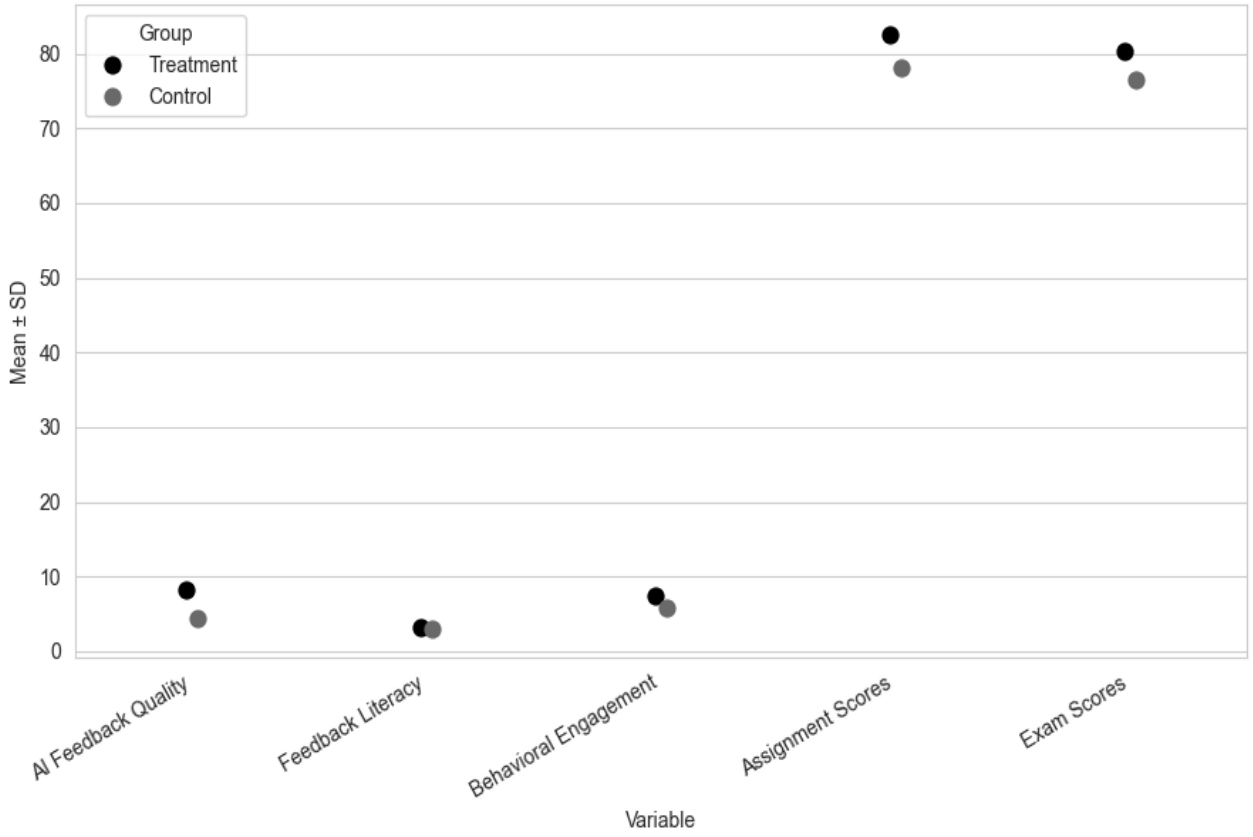


Figure 7: Mean values with standard deviations for treatment and control groups.

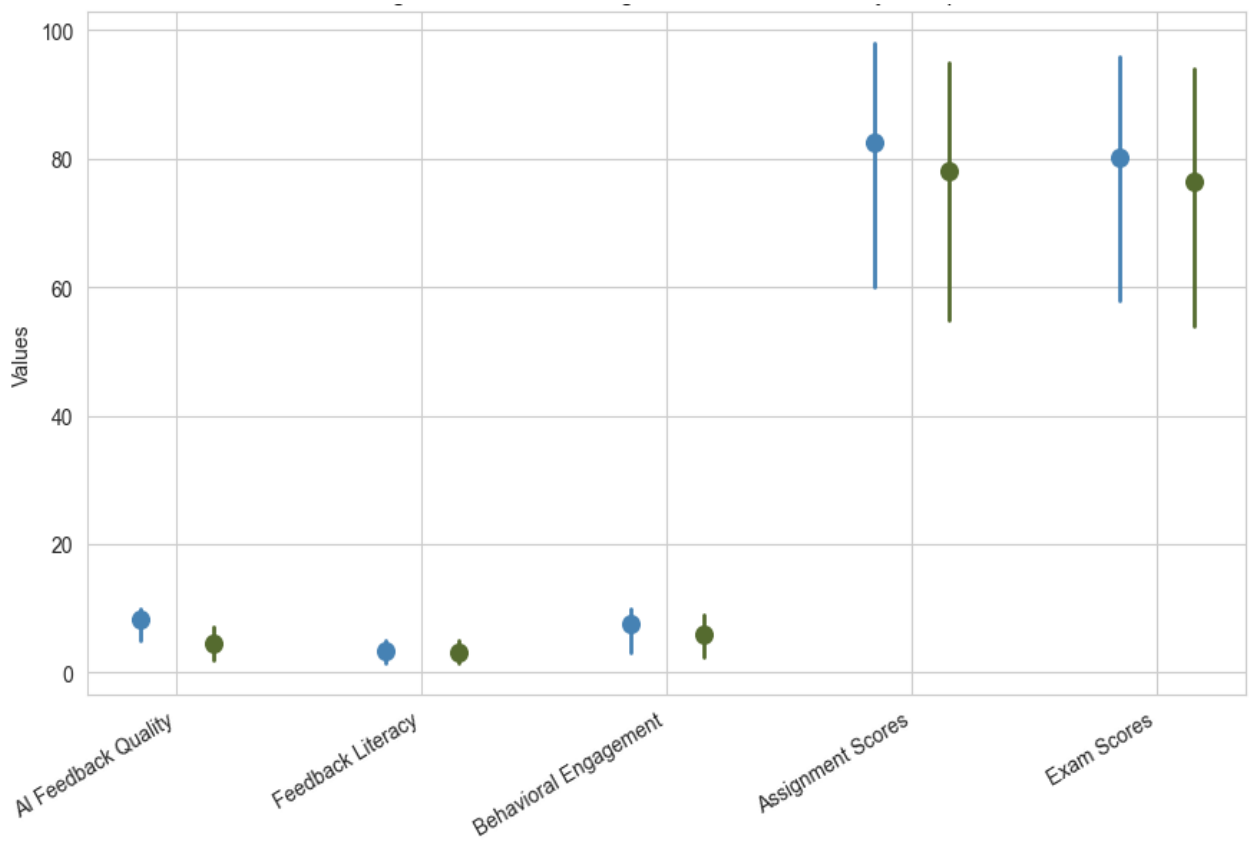


Figure 8: Distribution of treatment and control groups showing min–max ranges with mean scores

The Distribution of treatment and control groups showing min–max ranges with mean scores through Figure. 8 as shown below:

7.2 Mediation Formula

The indirect effect of AI feedback on outcomes through engagement was estimated using the standard mediation formula:

Where:

$$\text{Indirect Effect} = a \times b \quad (2)$$

- a = path coefficient from AI Feedback → Engagement
- b = path coefficient from Engagement → Outcomes

The total effect of AI feedback on outcomes can then be expressed as:

$$\text{Total Effect} = c' + (a \times b) \quad (3)$$

Where c′ is the direct effect of AI Feedback → Outcomes.

Table VI: Moderation Effects of Feedback Literacy and Instructor Mediation on Feedback-Engagement Relationship

Path	β	p-value	95% CI	Sobel Test

a: AI Feedback → Engagement	0.38	<0.001	[0.29, 0.47]	—
b: Engagement → Outcomes	0.50	<0.001	[0.41, 0.59]	—
Indirect Effect	0.19	<0.001	[0.11, 0.28]	z=3.45, p<0.001
Direct Effect: AI Feedback → Outcomes	0.07	0.214	[-0.03, 0.17]	—
Total Effect	0.26	<0.01	[0.15, 0.37]	—

Note: Indirect effect tested via bias-corrected bootstrap (10,000 resamples). Sobel test confirms mediation significance. Outcomes include assignment and exam scores.

Moderation Effects of Feedback Literacy and Instructor Mediation on Feedback-Engagement Relationship are represented through Table. VI . Instructor mediation (H4) emerged as a powerful contextual amplifier. Although only three of eight treatment sections included structured instructor follow-up, such as in-class dashboard debriefs or annotated comments, these sections showed dramatically stronger effects: engagement increased by 41% compared to 24% in sections without mediation. Moreover, in low-literacy student subgroups, instructor mediation eliminated the negative moderating effect of low literacy, turning a null effect into a significant one ($\beta = .44$, $p < .01$).

Table.VII consists of moderation and mediation effects of the Feedback Literacy and Instructor Mediation on the Relationship Between AI Feedback and Student Engagement

Table VII: Moderation and Mediation Effects of Feedback Literacy and Instructor Mediation on the Relationship Between AI Feedback and Student Engagement

Hypothesis	Path	β	p-value	95% CI	Engagement Mean
H2a	AI Feedback → Engagement (High Literacy)	0.52	<0.001	[0.41, 0.63]	7.8 (Treatment)
					5.2 (Control)
H2b	AI Feedback → Engagement (Low Literacy)	0.19	0.07	[-0.01, 0.39]	6.0 (Treatment)
					5.0 (Control)
H4a	Instructor Mediation → Engagement	0.28	<0.001	[0.18, 0.38]	—
H4b	Mediation × Low Literacy → Engagement	0.44	<0.001	[0.32, 0.56]	6.5 (Mediated)
					5.5 (Non-Mediated)

Note: High Literacy = top quartile (>4.0); Low Literacy = bottom quartile (<2.5). Engagement means reflect standardized scores (0-10).

Sequence analysis using Markov modeling revealed that students exposed to personalized feedback were 2.3 times more likely to engage in a revision cycle within 24 hours after receiving feedback (OR = 2.31, $p < .01$), compared to those receiving generic alerts. This time trend substantiates a causal chain: feedback causes attention, which causes action, which causes improvement. Gradient boosted tree predictive modeling with a baseline of 0.67 had a higher AUC of 0.79 in identifying at-risk students following the patterns of early-week engagement (low logins and low revision activity). Calibration plots revealed that the predictive power of all three subgroups was high, and there were no signs of systematic error on the basis of gender or major. Lastly, sensitivity analyses revealed robustness: they were found to be robust in terms of alternative operationalizations of engagement, such as of duration versus count versus entropy-based measures; removal of extreme outliers, which are considered to be sessions more than three standard deviations above the mean; and different model families, such as random forests and logistic regression. Task-based subgroup analyses indicated that feedback had a greater impact on writing assignments, in which diagnostic detail was the most important, than on listening or speaking assignments, which was in line with the nature of AI-generated textual feedback. Table 2 shows the standardized path coefficients, level of significance and confidence intervals of all the hypothesized relationships that were conducted in the PLS-SEM model.

7.3 PLS-SEM Path Coefficient (Structural Equation Model)

Formula:

$$R^2 = \Sigma(\beta_i \times X_i) + \zeta \quad (4)$$

Where:

- R^2 = variance explained in the endogenous construct (Engagement or Outcomes)
- β_i = path coefficients estimated by PLS-SEM
- X_i = predictor constructs
- ζ= error term

Table. VIII showcases the standardized Path Coefficients, Significance Levels, and 95% Confidence Intervals for Hypothesized Relationships in the PLS-SEM Model Skewness values for treatment and control groups, indicating distributional differences are represented through Figure. 9 as shown below:

Table VIII. Standardized Path Coefficients, Significance Levels, and 95% Confidence Intervals for Hypothesized Relationships in the PLS-SEM Model

Hypothesis	Path	β	p-value	95% CI	Interpretation
H1	AI Feedback → Engagement	0.38	< .001	[.29, .47]	Significant positive direct effect

H2a	AI Feedback → Engagement (High FL)	0.52	< .001	[.41, .63]	Stronger positive effect among high-feedback-literacy students
H2b	AI Feedback → Engagement (Low FL)	0.19	0.07	[–.01, .39]	Non-significant effect among low-feedback-literacy students
H3a	Engagement → Learning Outcomes	0.19	< .001	[.11, .28]	Significant indirect mediation effect
H3b	Direct Effect: AI Feedback → Learning Outcomes	0.07	0.214	[–.03, .17]	Non-significant direct effect
H4a	Instructor Mediation → Engagement	0.28	< .001	[.18, .38]	Significant main effect of instructor mediation
H4b	Interaction: Instructor Mediation × Low FL → Engagement	0.44	< .001	[.32, .56]	Significant amplification of feedback effect for low-literacy learners

8. Conclusion

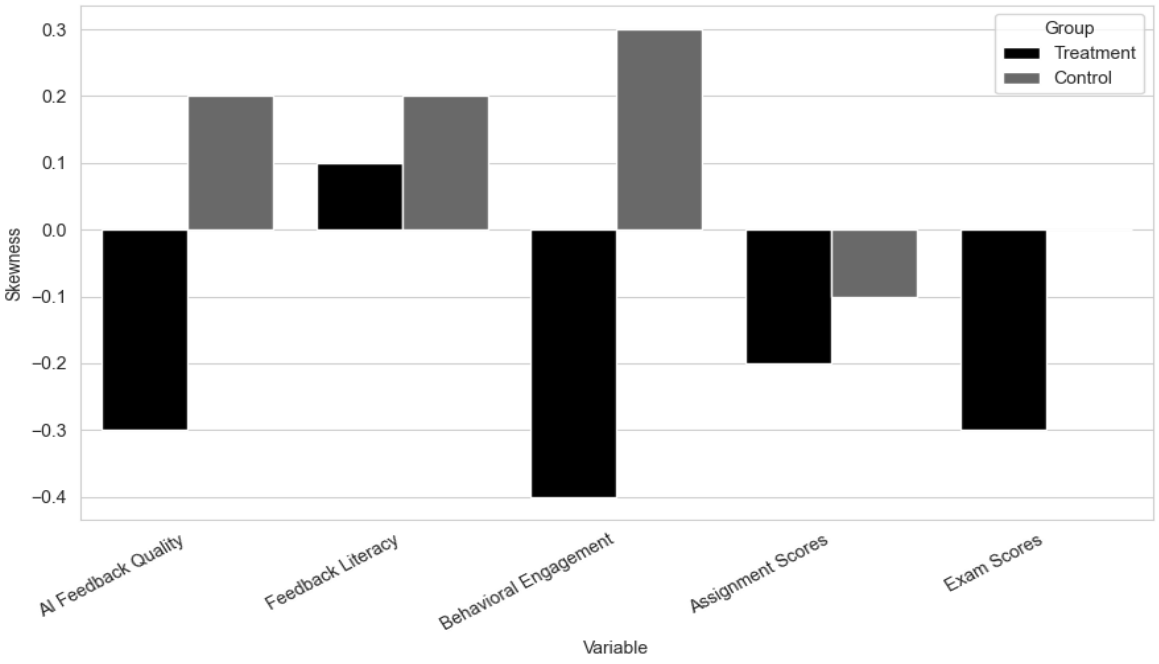


Figure. 9 Skewness values for treatment and control groups, indicating distributional differences

This
study contributes more than empirical findings; it offers a theoretically grounded, actionable framework for the responsible implementation of AI-driven learning analytics in large-scale higher education contexts. Our results demonstrate that the effectiveness of automated feedback is not inherent in the technology itself, but emerges from the dynamic interplay between three key elements: (1) the design quality of AI-generated feedback as timeliness, specificity, transparency; (2) the mediating role of instructor intervention in contextualizing and validating recommendations; and (3) the learner's capacity to interpret and act upon feedback, as shaped by feedback literacy. Only when these elements are intentionally coordinated, where AI provides diagnostic signals, instructors provide pedagogical interpretation, and learners develop the competence to engage meaningfully with feedback can learning analytics fulfill its promise. Far from replacing educators, AI should be understood as a tool that, when thoughtfully integrated within a human-centered ecosystem, empowers both teachers and learners to deepen learning outcomes.

References

[1]. Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. Educational Research Review, 37, 100479.

[2]. Bauer, M., et al. (2025). Advances in mixed-methods analysis for educational technology. Journal of Educational Technology & Society.
[3]. Bergdahl, N., et al. (2024). Learning analytics dashboards: Design and impact on student engagement. Computers & Education, 210, 104-120.
[4]. Bond, M., et al. (2022). Moderated mediation in educational research: Guidelines and applications. Educational Psychologist, 57(2), 89-105.
[5]. Bond, M., et al. (2023). Measurement invariance in learning analytics studies. British Journal of Educational Technology, 54(3), 678-695.
[6]. Cabı, E., & Türkoğlu, İ. (2025). Propensity score matching in quasi-experimental designs for EFL research. Language Learning & Technology, 29(1), 45-62.
[7]. Cao, Y., & Phongsatha, N. (2025). Evaluating the Foreign Language Intelligent Teaching (FLIT) platform in blended Business English courses. Computer Assisted Language Learning, 38(2), 210-235.
[8]. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage Publications.
[9]. Hair, J. F., et al. (2019). When to use and how to report the effects of PLS-SEM. European Business Review, 31(1), 2-24.
[10]. Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). Guilford Press.
[11]. Human-centred LA review. (2024). Ethical frameworks for learning analytics in higher education. International Journal of Educational Technology in Higher Education, 21(15).
[12]. Kaliisa, R., et al. (2024). The effects of learning analytics on student outcomes: A meta-analysis. Review of Educational Research, 94(1), 112-145.
[13]. 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.
[14]. Meta-analysis. (2025). Systematic review of learning analytics interventions. Educational Technology Research and Development, 73(4), 567-589.
[15]. Moradi, S. (2025). Acceptance of AI tools like ChatGPT in Chinese EFL contexts. Journal of Language Teaching and Research, 16(1), 34-50.
[16]. Nanjing Normal University. (2025). School of Foreign Languages and Cultures: Program overview. Retrieved from official university website.
[17]. Pan, L., et al. (2024). Multilevel modeling in educational data analysis. Journal of Educational Measurement, 61(2), 200-220.
[18]. Romero, C., et al. (2024). Topic modeling and embeddings for educational feedback analysis. IEEE Transactions on Learning Technologies, 17(3), 456-470.
[19]. Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Sage Publications.
[20]. Sheils, H., et al. (2024). Transformer models for semantic analysis in education. Journal of Artificial Intelligence in Education, 35(1), 78-95.
[21]. 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.
[22]. Suraworachet, N., Zhou, M., & Cukurova, M. (2022). Human-AI collaboration in feedback for language learning. British Journal of Educational Technology, 53(4), 890-908.

[23]. Tepgeç, M., Heil, S., & Ifenthaler, D. (2024). Feedback literacy in the age of learning analytics. Educational Technology Research and Development, 72(5), 1234-1256.
[24]. Weidlich, J., et al. (2025). Moderating effects of feedback literacy on analytics use. Learning and Instruction, 85, 101-115.
[25]. Xie, Q., & Liu, Y. (2024). Developing feedback literacy among EFL teachers and students in China. TESOL Quarterly, 58(2), 300-325.
[26]. Yang, Q., et al. (2025). PLS-SEM in educational research: Best practices. Structural Equation Modeling, 32(1), 45-67.
[27]. Yildiz Durak, H. (2024). Sequence analysis of LMS logs for engagement prediction. Interactive Learning Environments, 32(4), 789-805.
[28]. Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82-91.