

Academic engagement and artificial intelligence platform behaviors in grammar achievement

Wang Yadan¹, Soon Singh Bikar Singh¹, Connie Shin¹, Zheng Juncai², Zhang Qianqian²

¹Faculty of Education and Sports Studies, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia

²School of International Studies, Guangzhou Xinhua University, Guangzhou, China

Article Info

Article history:

Received Nov 23, 2025

Revised Mar 12, 2026

Accepted Mar 28, 2026

Keywords:

Academic engagement

Artificial intelligence

Grammar achievement

Higher education

Platform behaviors

ABSTRACT

This study is among the first to use archival institutional records to test the incremental validity of artificial intelligence platform behaviors (AI_index) in predicting grammar achievement (GA). Using data from 405 non-English-major freshmen enrolled in a compulsory grammar course at a private Chinese university, we examined whether AI_index predicts end-of-semester grammar exam performance beyond course-embedded behavioral academic engagement (AE_index). AE_index was derived from grade-book quizzes and class interactions, whereas AI_index was constructed from institutional platform logs capturing coursework completion and assigned video viewing. Indices were scaled to a 0–100 range, and GA was measured by a unified final exam. Descriptive statistics, correlations, and hierarchical regression analyses showed that AE_index was a small but significant predictor of exam performance, whereas AI_index was weak and non-significant and added no incremental predictive value beyond AE_index. Together, the two indices explained a modest proportion of variance in GA. These findings suggest that completion-based platform metrics are unlikely to reflect effortful learning unless platform tasks align with summative assessment demands (e.g., translation and proofreading). The findings caution against using completion-based AI metrics as high-stakes indicators without demonstrated task–assessment alignment.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Soon Singh Bikar Singh

Faculty of Education and Sports Studies, Universiti Malaysia Sabah

Kota Kinabalu 88400, Sabah, Malaysia

Email: soonbs@ums.edu.my

1. INTRODUCTION

The review is organized around three themes: i) AI in language education and learning analytics; ii) engagement–achievement links; and iii) grammar pedagogy and assessment alignment. Within each theme, evidence is synthesized at the construct level. The rapid development of artificial intelligence (AI) has a major impact on education. As highlighted in UNESCO policy guidance, both opportunities and risks exist, particularly regarding inclusion and equity in the deployment of AI in education [1]. Specifically, in language education, AI tools can provide real-time feedback [2], but their effectiveness hinges on educators' ability to integrate them into teaching practices; over-reliance on AI-generated corrections may result in passive learning and surface-level engagement [3]. In parallel to general developments, AI-mediated language instruction has been empirically examined with university English as a foreign language (EFL) learners in China [4]. Moving from individual instructional contexts to broader applications, recent work on integrating corpora and generative artificial intelligence (GenAI) within data-driven learning (DDL) highlights both new possibilities and limitations, including limited tracking in corpus-based approaches and potential inaccuracies

in GenAI outputs [5]. At the system level, information system platforms may support learning when course management and platform design (e.g., guidance and real-time feedback) align with learners' needs and course requirements. However, limited resources and adaptability may hinder academic success [6]. Accordingly, it is important to recognize that completion-based platform records may underrepresent effortful learning. Taken together, this literature suggests that platform logs often capture exposure or compliance, and their interpretability as learning indicators depends on task–assessment alignment.

Turning to the connection between engagement and achievement, students' academic engagement has been linked to outcomes such as test scores and grades in research [7]. Engagement is widely regarded as a multidimensional construct with behavioral, emotional, and cognitive components [8]. Behavioral engagement involves on-task attention, effort, and persistence; emotional engagement reflects task-facilitating emotions; and cognitive engagement involves deep learning and self-regulatory strategies [9]. In technology-enhanced EFL contexts, structured digital participation has been associated with improved language learning outcomes, including speaking performance in a Flipgrid-based program with Jordanian adolescent EFL learners [10] and grammar learning outcomes in gamified online English courses in Thai higher education [11]. Meta-analytic evidence further supports positive engagement–achievement associations in general education ($r=0.27$) [12] and second-language learning ($r=0.23$), with stronger associations when engagement is measured over a learning unit rather than as a single task or general measure [13]. Accordingly, the present study focuses on course-embedded behavioral academic engagement (AE_index) captured in institutional records, rather than self-reported emotional or cognitive engagement, to align with the analytics purpose of modeling variance in the same course's summative grammar achievement (GA) and AI platform behaviors (AI_index).

Grammar remains a central issue in L2 pedagogy, and early form-focused work provides a knowledge base for subsequent meaning-focused learning [14]. In grammar assessment, test tasks and score interpretations should be guided by the test purpose and reflect relevant target language use (TLU) tasks (i.e., test authenticity) [15]. Evidence from a systematic review of online higher education shows that self-regulated learning (SRL) strategies—especially time management, metacognition, and effort regulation—are significantly but weakly associated with academic achievement [16], and SRL is commonly framed as encompassing cognitive, metacognitive, behavioral, motivational, and emotional/affective processes [17]. These findings suggest that integrating traditional and AI-assisted strategies could support language learning [18]. From a validity perspective, response formats and task demands are central to what grammar scores can meaningfully represent.

Despite these insights, the problem remains that evidence is still limited on the joint and incremental contributions of course-embedded behavioral academic engagement and AI platform behaviors to summative GA in private Chinese universities. Addressing this gap can inform the pedagogical integration of AI-supported resources in grammar courses. The key empirical issue is whether AI platform behavioral records (e.g., completion-based traces) provide incremental predictive value for grammar exam performance beyond course-embedded behavioral academic engagement derived from institutional course records. This issue is especially salient when platform traces primarily reflect task completion, whereas the summative grammar exam targets specific response formats and elicited abilities. Because expected response types are defined by the task input and provide the basis for score-based inferences about grammatical knowledge, mismatches in task/response formats can weaken the interpretability of platform traces as predictors of exam outcomes [15].

In this course, platform logs capture completion-oriented, auto-scored practice, while the exam requires constructed-response proofreading and translation; this mismatch in format may weaken the relationship between AI platform behaviors and GA. The present study examines whether course-embedded behavioral academic engagement and AI platform behaviors predict GA in a compulsory grammar course at a private Chinese university. Figure 1 illustrates the conceptual model, specifying AE_index and AI_index as concurrent predictors of GA, with particular attention to the incremental contribution of AI_index beyond AE_index. GA is measured by the percentage score on a unified end-of-semester exam; the exam blueprint and item formats are reported in the methods section.

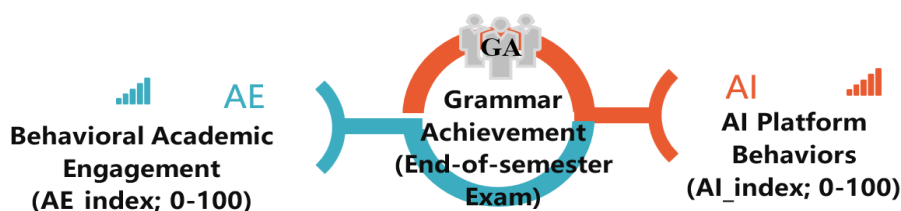


Figure 1. Conceptual model of AE_index and AI_index predicting GA

The study contributes by: i) integrating institutional grade-book records with platform traces in an authentic course setting; ii) operationalizing AE_index and AI_index use as behavioral composites for predictive validity testing; and iii) offering practical implications for aligning platform activities with assessed response formats to support sustained, form-focused grammar practice. Accordingly, two analytic questions were examined:

- Do AE_index and AI_index jointly predict GA?
- Does AI_index explain additional variance in GA beyond AE_index (incremental validity)?

2. METHOD

2.1. Research design and participants

This study used a quantitative correlational design with archival institutional records. In this course, the institutionally mandated weighting scheme was preset in the AI platform; after continuous-assessment component scores were recorded, the platform automatically computed students' usual performance. Once the instructors finished marking the end-of-semester exam, they entered the scores into the system. The platform then calculated the overall course evaluation. All variables came from archival institutional records, including the official grade book and the institutional AI platform.

Participants were 405 non-English majors in a required English grammar course at a private university in southern China. The data came from six intact classes, all following a common syllabus with a centrally coordinated teaching team. Each class completed the same continuous assessment components and took the same end-of-semester exam. The course used a blended format: face-to-face instruction plus required platform-based tasks. Continuous assessment, called "usual performance", contributed 60% of the course grade. The end-of-semester exam contributed the remaining 40%.

2.2. Measures

All variables were drawn from behavioral records in the course grade book and AI platform logs. AE_index and AI_index were converted to a 0–100 scale for analysis, while component scores remained on their original scales. Figure 2 shows a breakdown of the usual performance (continuous assessment; 60%) recorded on the platform. It also illustrates how the system aggregates component scores. AE_index and AI_index were defined as formative composites of course-embedded behavioral components. Internal consistency indices (α/ω) are reported descriptively, not as scale reliability. Low internal consistency values are therefore expected and do not indicate measurement deficiency for these formative indicators. All records were anonymized and used with institutional approval.

Student ID anonymized	Class Attendance (10%)	Quiz (30%)					Class interaction (10%) : answering questions 5% +activity participation 5%			coursework (30%) : Chaoxing 15%+We Learn platform 15%					Platform-tracked video viewing completion (4 units) (20%)	Usual performance (0-100; 60%)	End-of- semester exam (0-100; 40%)	Overall evaluation (100%)
		1	2	3	4	SUM	Questions (0-5)	Activity (0-5)	SUM	1	2	3	4	SUM				
Illustrative	10	75	93	90	95	26.5	3.4	1.7	5.1	98	96	92	96	28.7	15.1	85.4	83	84.4

Figure 2. Anonymized screenshot of the institutional grade report showing assessment components and weights used to derive AE_index and AI_index (illustrative values only)

AE_index was calculated from two course-embedded behavioral components in Figure 2: quizzes (0–30) and class interaction (0–10). Class interaction included answering questions (0–5) and participation in activities (0–5). Attendance accounted for 10% of the official scheme but was excluded from the AE_index because nearly all students had perfect attendance. This resulted in little variance and did not help the analysis of individual differences. Attendance remained part of the official grading. Quizzes and class interaction were summed (0–40) and converted to a 0–100 score. AE_index was specified as a formative composite because its components measure distinct aspects of engagement. Internal consistency is reported descriptively ($\alpha=0.550$; $\omega_{total}=0.634$). Collinearity among components was negligible (all variance inflation factor (VIFs) ≤ 1.443).

AI_index was computed from the platform-logged weighted components in Figure 2: coursework (0–30) and platform-tracked video viewing completion (0–20). The summed score (0–50) was converted to a 0–100 score and is interpreted as a course-embedded indicator of platform participation rather than depth of cognitive processing. Coursework consisted of unit-level grammar drills, auto-scored quizzes, and short

practice tests. Video viewing captured completion of assigned instructional videos and embedded checks. AI index was treated as a formative composite because coursework performance and video-viewing completion capture conceptually distinct facets of platform behavior that jointly define the indicator rather than reflecting a single latent trait. Internal consistency is reported descriptively ($\alpha=0.283$; $\omega_{\text{total}}=0.285$).

GA was measured using the official end-of-semester exam (0–100). Test score interpretation and intended use followed contemporary validity theory, which focuses on the appropriateness of inferences and uses from assessment outcomes [19]. The closed-book, 90-minute exam was given to all classes simultaneously. It was centrally prepared, reviewed by the teaching team, and graded with standardized rubrics. The exam was designed for institutional summative use and matched the syllabus objectives as specified in the institutional syllabus [20]. It covered core grammatical areas (e.g., tense/aspect, agreement, clause structure) using standard formats: multiple-choice, error identification or correction, and Chinese–English translation requiring grammatically correct responses. Constructed-response formats were included to assess language use, not just recognition. The methods section reports the exam blueprint and item formats for transparency. Table 1 provides the blueprint and formats. The archival dataset lacked item-level responses and rater data, so reliability indices could not be computed.

Table 1. Blueprint of the end-of-semester exam of the compulsory grammar course

Part	Section	Item type and description	No. of items	Points per item	Section points	Weight (%)
I	Vocabulary and structure	Multiple-choice questions assessing grammatical forms and vocabulary use	20	1.5	30	30
II	Proofreading	Error identification/correction in short sentences or passages	10	1.5	15	15
III	Sentence translation	Chinese–English sentence translation (focus on grammatical accuracy)	5	6.0	30	30
IV	Paragraph translation	Chinese–English paragraph translation (one passage; 5 sentences)	1 passage	25.0 (total)	25	25
Total					100	100

2.3. Data collection and analysis

At the end of the semester, the teaching team exported continuous assessment scores from the official grade book and platform task records from the institutional AI platform. Student identifiers were removed and replaced with anonymized codes before analysis, and exported values were checked for permissible ranges and consistency with the official weighting scheme. Of the 410 original course records, five were excluded due to incomplete GA records associated with withdrawal or a leave of absence, yielding a final analytic sample of $N=405$. Descriptive statistics (mean, standard deviation, minimum, maximum) were computed for AE_index, AI_index, and GA, and score distributions were inspected for plausible ranges and potential outliers. Pearson correlation coefficients were then calculated to examine bivariate relationships among academic engagement, AI platform behaviors, and GA. Third, a standard multiple regression model was specified with GA as the dependent variable and both AE_index and AI_index as predictors entered simultaneously. To directly assess incremental validity, hierarchical regression was conducted in two steps: in Step 1, only AE_index was entered as a predictor; in Step 2, both AE_index and AI_index were included. The change in explained variance (ΔR^2) and F-change statistics were examined after Step 2 to determine if the addition of AI_index significantly improved model prediction. The regression model can be expressed as in (1):

$$\text{Grammar achievement} = \beta_0 + \beta_1(\text{AE_index}) + \beta_2(\text{AI_index}) + \varepsilon \quad (1)$$

Assumptions of multiple regression were evaluated using residual plots and normal probability (P–P) plots for residual normality [21]. No major violations were observed. Multicollinearity was checked using tolerance/VIF and was negligible ($\text{VIF}_{\text{AE_index}}=1.188$; $\text{VIF}_{\text{AI_index}}=1.188$). Statistical significance was evaluated at the 0.05 level. Both unstandardized coefficients (B) and standardized coefficients (β) were examined. Model R^2 was also reviewed. As a robustness check, the regression model was re-estimated with Attendance as an additional covariate. Results remained substantively unchanged, consistent with the minimal attendance variance in this cohort.

2.4. Ethical considerations

Secondary use of course-related records was approved by the university's academic affairs office on 13 September 2025. Students were informed at the beginning of the semester that de-identified course and platform records might be used for research, and the approving office granted a waiver of written informed

consent for this minimal-risk secondary analysis. Only variables required for the present study were extracted, and identifiers were removed prior to analysis. The de-identified dataset was stored on an access-controlled, password-protected institutional drive restricted to the research team and was not shared externally (including with the platform provider) in identifiable form. Files were stored and transferred only through secure institutional channels. The dataset will be retained for three years and then securely deleted. Because some measures involved teacher scoring (e.g., questions and activity participation), predefined scoring rules and shared guidance within the teaching team were used to reduce subjectivity; residual scoring bias and power dynamics are acknowledged as limitations. The research team accessed only de-identified institutional exports provided by the academic affairs office and had no access to provider-side logs beyond these exports. This de-identification and access-control protocol improves transparency and reduces the risk that institutional analytics or completion indicators could be used to differentially treat identifiable students or inform high-stakes decisions without construct-relevant validation.

3. RESULTS AND DISCUSSION

3.1. Data screening and descriptive statistics

All analyses used the final analytic sample (N=405). Visual inspection confirmed that all remaining values for continuous assessment components, platform tasks, and the end-of-semester exam fell within plausible ranges. No extreme outliers were identified. Outliers were screened using boxplots and z-scores ($|z| > 3.29$); none met the criterion.

Table 2 presents descriptive statistics for AE_index, AI_index, and GA. AE_index and AI_index are interpreted as course-embedded behavioral composites derived from institutional records, not as comprehensive measures of multidimensional engagement. Descriptively, AE_index (M=70.37, SD=21.84) and AI_index (M=80.55, SD=16.32) suggest moderately high AE_index and relatively high platform-tracked task completion/activity. GA was also relatively high on average (M=77.82, SD=12.93). The observed ranges support subsequent correlational and regression analyses.

Table 2. Descriptive statistics for AE_index, AI_index, and GA (N=405)

Variable	Mean	SD	Min	Max
Behavioral academic engagement (AE_index)	70.37	21.84	1.15	100.00
AI platform behaviors (AI_index)	80.55	16.32	0.55	99.77
GA	77.82	12.93	7.50	98.50

3.2. Correlations among AE_index, AI_index, and grammar achievement

Pearson correlations among AE_index, AI_index, and GA are reported in Table 3. AE_index and AI_index were moderately and positively correlated ($r=0.398$, $p<0.001$). This indicates that students who obtained higher scores on course-embedded behavioral components (quizzes and class interaction) also tended to show higher platform-tracked coursework performance and online resource completion. This pattern may indicate that students who engage more consistently in one course component also do so in other required components. GA showed a small but statistically significant positive association with AE_index ($r=0.221$, $p<0.001$). This suggests that higher AE_index was associated with higher end-of-semester exam scores. The magnitude of this association is broadly consistent with meta-analytic evidence showing modest positive engagement–achievement correlations (e.g., $r \approx 0.27$ in general education and $r \approx 0.23$ in L2 learning) [12], [13]. By contrast, the association between AI_index and GA was positive but weak, and not statistically significant ($r=0.065$, $p=0.188$). AI_index primarily reflects platform-logged activity/completion rather than depth of processing. The null association is interpreted as an operationalization/measurement consideration, not as a causal claim.

Table 3. Pearson correlations among AE_index, AI_index, and GA (N=405)

Variable	1	2	3
Behavioral academic engagement (AE_index)	1.00		
AI platform behaviors (AI_index)	0.398***	1.00	
GA	0.221***	0.065	1.00

Note: Pearson correlations are reported. *** $p<0.001$.

3.3. Regression analysis predicting grammar achievement

To address whether AE_index and AI_index jointly predict GA and whether AI_index adds incremental predictive value beyond AE_index, multiple regression and hierarchical regression were conducted. Step 1 included AE_index, and Step 2 added AI_index. Analyses used R^2 , ΔR^2 , and the F-change test. The overall model was statistically significant ($R^2=0.049$, Adjusted $R^2=0.045$, $F(2, 402)=10.418$, $p<0.001$). This indicates that the two course-embedded behavioral predictors together explained approximately 4.9% of the variance in GA. Although modest, this variance explained is meaningful in a real-course setting, where achievement is multifactorial and influenced by multiple learner and contextual factors. In the hierarchical test, AE_index accounted for 4.87% of the variance ($R^2=0.04869$, $F(1, 403)=20.63$, $p=7.39\times 10^{-6}$). Adding AI_index produced a negligible and non-significant increase ($R^2=0.04928$; $\Delta R^2=0.00059$; F-change (1, 402)=0.250, $p=0.618$). This pattern indicates limited incremental validity of AI_index for predicting exam performance beyond AE_index. Table 4 presents the regression coefficients. AE_index was a significant positive predictor of GA ($B=0.137$, $SE=0.031$, $\beta=0.231$, $t=4.362$, $p=1.6\times 10^{-5}$, 95% CI [0.075, 0.199]).

Table 4. Multiple regression predicting GA from AE_index and AI_index (N=405)

Predictor	B	SE	Std. β	t	p
Behavioral academic engagement (AE_index)	0.137	0.031	0.231	4.362	<0.001
AI platform behaviors (AI_index)	-0.021	0.042	-0.026	-0.500	0.618

Model statistics: $R^2=0.0493$, $F(2, 402)=10.418$, $p<0.001$. Coefficients correspond to Step 2 (full model)

Holding AI_index constant, a one-point increase in AE_index (0–100 scale) corresponded to an estimated 0.14-point increase in grammar exam score. To avoid construct overreach, interpretations are restricted to AE_index (quizzes and class interaction) rather than multidimensional engagement. By contrast, AI_index did not significantly predict GA after controlling for AE_index ($B=-0.021$, $SE=0.042$, $\beta=-0.026$, $t=-0.500$, $p=0.618$, 95% CI [-0.104, 0.062]). Regression assumptions were checked using residual diagnostics (including normal probability P–P plots) and collinearity statistics (tolerance/VIF). No major violations were indicated. Predictor VIF values were low (both=1.188), indicating negligible multicollinearity. Effect sizes were small for AE_index (Cohen's $f^2=0.047$) and negligible for AI_index ($f^2=0.001$). Figure 3 visualizes the standardized paths (β) and model R^2 (N=405).

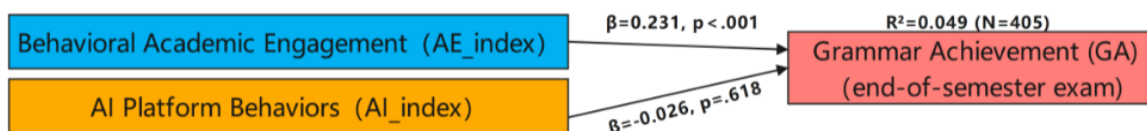


Figure 3. Regression path diagram predicting GA from AE_index and AI_index (standardized coefficients)

3.4. Integrated discussion

Consistent with the results, course-embedded behavioral academic engagement (AE_index) showed incremental predictive value for GA, whereas AI platform behaviors (AI_index) did not add explanatory power beyond AE_index, suggesting limited criterion relevance of completion-based platform indicators for this end-of-semester exam. Overall, the two indices accounted for a modest proportion of variance in exam performance, indicating that observable engagement contributes to grammar outcomes but does not fully explain individual differences in achievement. Why did AE_index predict GA? The association aligns with behavioral engagement perspectives in which participation, effort, and persistence are core indicators of engagement that consistently predict learning outcomes and achievement [7], [9]. In this course, quizzes and structured interaction likely created repeated practice–feedback opportunities that supported learners' error detection and self-correction, consistent with form-focused grammar learning processes [14]. In opportunity-to-learn terms, these behaviors provide repeated, feedback-supported practice on assessable forms that is closer to the exam's constructed-response demands, making them temporally and cognitively proximal to the summative assessment. Supportive course contexts may facilitate higher-quality motivation and performance by supporting competence, autonomy, and relatedness, while performance outcomes also reflect both learner differences and contextual supports [22].

From an assessment validity perspective, AI_index's non-significant contribution more likely reflects construct misalignment between the platform logs and the grammar exam than null effects of

AI-supported practice. AI_index primarily captured completion and auto-scored exposure, whereas the exam emphasized constructed-response proofreading and translation; such response-format differences can add format-related noise and threaten the validity of linking such indicators to constructed-response scores [15]. This also underscores that completion traces may miss instructionally meaningful supports. Structure involves clear expectations and goals, consistent rules and guidelines, and informational supports for engagement and efficacy feedback [23]. Accordingly, completion-based analytics should not be used as learning proxies without validation evidence that the indicator is construct-relevant to the intended assessment inferences and outcomes [24]. Technology-enhanced learning validity syntheses stress interpreting evidence in relation to what is measured, how it is used, and whether indicators support the intended validity claims and uses [25]. From a closed-loop learning analytics perspective, trace indicators should inform iterative refinement of tasks, supports, and feedback to improve construct alignment and instructional impact [26]. Trace-validity scholarship similarly cautions that trace indicators require theory-driven alignment and validation, rather than assuming logged actions directly represent learning processes [27]. Trace-based self-regulation work shows that digital traces can diverge from self-reports and are often only weakly aligned, warranting cautious interpretation when relating traces to learning outcomes [28].

Engagement–achievement meta-analyses typically show small-to-moderate correlations that vary by engagement operationalization and study characteristics (e.g., measurement choices and time frame) [12]. In L2 learning, the engagement–achievement link is likewise modest (overall $r \approx 0.23$) and varies by learning context/time scale [13]. Thus, effect sizes should be interpreted in light of the specific engagement dimension and indicator captured by the traces, rather than as a generic “engagement” effect [29]. More generally, effect-size interpretation in education is context-sensitive, and even small effects may be meaningful when findings have scalable, cost-effective instructional implications [30].

Practical implications operate at three levels. Instructionally, frequent low-stakes quizzes and structured interaction should remain central in grammar teaching, sustaining active processing and feedback for error detection and correction in constrained production tasks. At the platform-task level, institutions should redesign AI-mediated tasks to better match exam-relevant response formats (e.g., constructed-response error detection and controlled translation) and to emphasize feedback uptake rather than completion. From a task–technology fit perspective, effective use depends on alignment among technological functionality, learner capabilities, and task requirements [31]. At the policy level, the findings caution against using completion-based AI metrics as high-stakes learning indicators without documented construct alignment and validity evidence linking traces to assessment outcomes; learning analytics should be validated for monitoring and accountability purposes, as dashboard-based predictions are not necessarily translated into pedagogical actions [32]. Limitations include reliance on archival indicators that under-represent cognitive/strategic engagement [33], lack of multi-source assessment/process evidence for finer-grained alignment analyses [34], and single-institution scope. Future research should combine cross-institutional replication with richer traces (e.g., time-on-task and feedback-uptake patterns) and explicit platform–exam alignment mapping to clarify when AI-supported behaviors predict GA.

4. CONCLUSION

Using archival institutional records from 405 non–English-major undergraduates, this study found that course-embedded behavioral academic engagement was a small but significant predictor of grammar exam performance, whereas AI platform behaviors were weak and non-significant and provided no incremental predictive value. The overall engagement–achievement association was modest. Completion-based platform metrics may have limited interpretability for grammar achievement unless platform tasks align with the exam’s response formats and task demands; improving task–exam alignment may enhance the interpretability of score-based inferences. Institutions should therefore treat completion-based platform indicators as participation metrics rather than high-stakes proxies in the absence of construct alignment and validity evidence. Future research should extend this work through longitudinal designs, cross-institutional replication, item-level exam analysis, and the integration of motivational, self-regulatory, and AI literacy measures.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Wang Yadan	✓	✓	✓		✓	✓		✓	✓	✓	✓			✓
Soon Singh Bikar Singh		✓		✓						✓		✓	✓	
Connie Shin					✓					✓		✓		
Zheng Juncai						✓	✓			✓				
Zhang Qianqian						✓	✓			✓				

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data were derived from internal course records of Guangzhou Xinhua University and are available from the corresponding author, [SSBS], upon reasonable request.




REFERENCES

- [1] UNESCO, *AI and Education: Guidance for Policy-makers*. Paris: UNESCO, 2021, doi: 10.54675/PCSP7350.
- [2] L. Wei, "Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning," *Frontiers in Psychology*, vol. 14, p. 1261955, Nov. 2023, doi: 10.3389/fpsyg.2023.1261955.
- [3] J. J. Jaramillo, A. Chiappe, and F. S. Delgado, "From struggle to mastery: AI-powered writing skills in ESL education," *Applied Sciences*, vol. 15, no. 14, p. 8079, Jul. 2025, doi: 10.3390/app15148079.
- [4] A. Mohebbi, "Enabling learner independence and self-regulation in language education using AI tools: a systematic review," *Cogent Education*, vol. 12, no. 1, p. 2433814, Dec. 2025, doi: 10.1080/2331186X.2024.2433814.
- [5] C. Lu, "AI-generated corpus learning and EFL learners' learning of grammatical structures, lexical bundles, and willingness to write," *PLOS One*, vol. 20, no. 7, p. e0321544, Jul. 2025, doi: 10.1371/journal.pone.0321544.
- [6] C. Yang and S. S. B. Singh, "User experience in information system platforms: a study on learning styles and academic challenges," *Journal of Internet Services and Information Security*, vol. 14, no. 4, pp. 209–223, Nov. 2024, doi: 10.58346/JISIS.2024.14.012.
- [7] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: potential of the concept, state of the evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, Mar. 2004, doi: 10.3102/00346543074001059.
- [8] J. A. Fredricks, M. Filsecker, and M. A. Lawson, "Student engagement, context, and adjustment: addressing definitional, measurement, and methodological issues," *Learning and Instruction*, vol. 43, pp. 1–4, Jun. 2016, doi: 10.1016/j.learninstruc.2016.02.002.
- [9] J. Reeve, "A self-determination theory perspective on student engagement," in *Handbook of Research on Student Engagement*, 1st ed., S. L. Christenson, A. L. Reschly, and C. Wylie, Eds., Boston, MA: Springer US, 2012, pp. 149–172, doi: 10.1007/978-1-4614-2018-7_7.
- [10] W. M. Zghoul and R. F. Bataineh, "Flipgrid: unlocking the English speaking potential of Jordanian adolescent EFL learners," *Journal of Information Technology Education: Innovations in Practice*, vol. 23, p. 17, 2024, doi: 10.28945/5407.
- [11] B. Waluyo, S. Phanrangsee, and W. Whanchit, "Gamified grammar learning in online English courses in Thai higher education," *Online Journal of Communication and Media Technologies*, vol. 13, no. 4, p. e202354, Oct. 2023, doi: 10.30935/ojcm/13752.
- [12] H. Lei, Y. Cui, and W. Zhou, "Relationships between student engagement and academic achievement: a meta-analysis," *Social Behavior and Personality: An International Journal*, vol. 46, no. 3, pp. 517–528, Mar. 2018, doi: 10.2224/sbp.7054.
- [13] A. Okunuki and Y. Kashimura, "Student engagement and academic achievement in second language learning: a meta-analysis," *JACET Journal*, vol. 68, pp. 71–90, 2024.
- [14] R. Ellis, "Current issues in the teaching of grammar: an SLA perspective," *TESOL Quarterly*, vol. 40, no. 1, pp. 83–107, Mar. 2006, doi: 10.2307/40264512.
- [15] J. E. Purpura, "Assessing grammar," in *The Companion to Language Assessment*, A. J. Kunnan, Ed., Hoboken, NJ: Wiley-Blackwell, 2013, pp. 100–124, doi: 10.1002/9781118411360.wbcla147.
- [16] J. Broadbent and W. L. Poon, "Self-regulated learning strategies & academic achievement in online higher education learning environments: a systematic review," *The Internet and Higher Education*, vol. 27, pp. 1–13, Oct. 2015, doi: 10.1016/j.iheduc.2015.04.007.
- [17] E. Panadero, "A review of self-regulated learning: six models and four directions for research," *Frontiers in Psychology*, vol. 8, p. 422, Apr. 2017, doi: 10.3389/fpsyg.2017.00422.
- [18] N. Jomaa, B. A. Mudhsh, and K. AlGhafri, "Traditional strategies and AI-integrated strategies in learning English among EFL Omani students," *Jurnal Arbitrer*, vol. 12, no. 3, pp. 368–382, Sep. 2025, doi: 10.25077/ar.12.3.368-382.2025.




- [19] S. Shaw and V. Crisp, "Tracing the evolution of validity in educational measurement: past issues and contemporary challenges," *Research Matters*, no. 11, pp. 14–19, 2011.
- [20] Guangzhou Xinhua University, "University foreign languages—grammar teaching syllabus compilation (Revised 2024)," Unpublished institutional document, 2024.
- [21] R. W. Osborne and E. Waters, "Four assumptions of multiple regression that researchers should always test," *Practical Assessment, Research and Evaluation*, vol. 8, no. 2, pp. 1–9, 2002.
- [22] E. L. Deci and R. M. Ryan, "The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior," *Psychological Inquiry*, vol. 11, no. 4, pp. 227–268, Oct. 2000, doi: 10.1207/S15327965PLI1104_01.
- [23] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions," *Contemporary Educational Psychology*, vol. 61, p. 101860, Apr. 2020, doi: 10.1016/j.cedpsych.2020.101860.
- [24] P. G. de Barba, E. A. Oliveira, and N. English, "Development and validation of a learning analytics rubric for self-regulated learning," *Educational Technology Research and Development*, vol. 73, no. 5, pp. 3223–3245, Oct. 2025, doi: 10.1007/s11423-025-10521-x.
- [25] M. van Haastrecht, M. Haas, M. Brinkhuis, and M. Spruit, "Understanding validity criteria in technology-enhanced learning: a systematic literature review," *Computers & Education*, vol. 220, p. 105128, Oct. 2024, doi: 10.1016/j.compedu.2024.105128.
- [26] M. Sailer, M. Ninaus, S. E. Huber, E. Bauer, and S. Greiff, "The end is the beginning is the end: the closed-loop learning analytics framework," *Computers in Human Behavior*, vol. 158, p. 108305, Sep. 2024, doi: 10.1016/j.chb.2024.108305.
- [27] Y. Fan *et al.*, "Towards investigating the validity of measurement of self-regulated learning based on trace data," *Metacognition and Learning*, vol. 17, no. 3, pp. 949–987, Dec. 2022, doi: 10.1007/s11409-022-09291-1.
- [28] F. Han and R. A. Ellis, "Self-reported and digital-trace measures of computer science students' self-regulated learning in blended course designs," *Education and Information Technologies*, vol. 28, no. 10, pp. 13253–13268, Oct. 2023, doi: 10.1007/s10639-023-11698-5.
- [29] N. Bergdahl, M. Bond, J. Sjöberg, M. Dougherty, and E. Oxley, "Unpacking student engagement in higher education learning analytics: a systematic review," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, p. 63, Dec. 2024, doi: 10.1186/s41239-024-00493-y.
- [30] S. Tobler, "Context matters: interpreting effect sizes in education meaningfully," *MethodsX*, vol. 13, p. 103023, Dec. 2024, doi: 10.1016/j.mex.2024.103023.
- [31] W. M. Al-Rahmi *et al.*, "Integrating communication and task–technology fit theories: the adoption of digital media in learning," *Sustainability*, vol. 15, no. 10, p. 8144, May 2023, doi: 10.3390/su15108144.
- [32] I. Masiello, Z. (Artemis) Mohseni, F. Palma, S. Nordmark, H. Augustsson, and R. Rundquist, "A current overview of the use of learning analytics dashboards," *Education Sciences*, vol. 14, no. 1, p. 82, Jan. 2024, doi: 10.3390/educsci14010082.
- [33] E. Mazzullo, O. Bulut, T. Wongvorachan, and B. Tan, "Learning analytics in the era of large language models," *Analytics*, vol. 2, no. 4, pp. 877–898, Nov. 2023, doi: 10.3390/analytics2040046.
- [34] P. Kannan and D. Zapata-Rivera, "Facilitating the use of data from multiple sources for formative learning in the context of digital assessments: informing the design and development of learning analytic dashboards," *Frontiers in Education*, vol. 7, p. 913594, Jun. 2022, doi: 10.3389/educ.2022.913594.

BIOGRAPHIES OF AUTHORS






Wang Yadan    is a Ph.D. candidate at the Faculty of Education and Sports Studies, Universiti Malaysia Sabah (UMS), Malaysia, and an assistant professor at Guangzhou Xinhua University, China. She received her M.A. in Foreign Linguistics and Applied Linguistics from Henan University, China. Her research interests include technology-enhanced English instruction and learning analytics in EFL contexts. She has published journal articles, including an EI-indexed paper, and has led four provincial/municipal research projects while participating in more than five additional funded projects. She can be contacted at email: wangyadan2025@gmail.com.






Soon Singh Bikar Singh    is the dean of the Faculty of Education and Sports Studies at Universiti Malaysia Sabah (UMS), Malaysia, and a Ph.D. supervisor. He holds a Ph.D. in Geography Education from Macquarie University, Sydney, an M.Ed. in Educational Psychology, and a B.A. (Hons) in Geography from the University of Malaya. His main research interests are in educational psychology, geography education, and curriculum studies, with particular attention to mixed methods research, the use of geographical information systems (GIS) in teaching. He has published book chapters and journal articles on GIS integration and geography skills in Malaysian schools and is active in professional associations. He can be contacted at email: soonbs@ums.edu.my.






Connie Shin    is an academic at Universiti Malaysia Sabah (UMS), Malaysia, and a PhD co-supervisor. She holds a PhD in Curriculum and Instruction from UMS and currently serves as the Head of the Rural Education Research Unit. She is also a Malaysian Qualifications Agency (MQA) auditor. Her main research interests include rural education, early childhood education, educational management and leadership, teaching and learning, STEM education in early childhood education, and intervention design and evaluation. She has contributed to curriculum development, teacher education, and educational innovation, particularly for rural and underserved communities. She can be contacted at email: conniecompok@ums.edu.my.



Zheng Juncai    is a Chinese–English translator and localization specialist based in Shenzhen, China. He holds an MA in Translation and Interpreting and is CATTI Level 2 certified. Since 2013, he has worked as a part-time Localization Engineer at Readdle Inc., focusing on UI localization, terminology management, MTPE, and cross-platform consistency for iOS and macOS products. He also works as a freelance translator in life sciences and game localization, and has contributed to MTPE/LQA and AI data projects with global platforms. He can be contacted at email: ijayci@gmail.com.



Zhang Qianqian    is a teaching assistant at Guangzhou Xinhua University, China. She holds a master's degree in Foreign Linguistics and Applied Linguistics. Her research interests focus on the psychological development mechanisms of ESP teachers, students' learning motivation, and the development and reform of college foreign language curricula. She has published several single-authored articles in national and international journals and developed the ESP course. She has led two national research projects and participated in three additional provincial-and university-level projects. She can be contacted at email: z1246919161@gmail.com.