AI-Mediated English Language and Literature Education: Digital Literacy, Cultural Identity, Student Engagement and the Moderating Role of Architectural Learning Environments

[1]Uzma Shaheen, [2]Mariam Saleem,[3]Ammara Afzaal,[4]Omar M. Shafiy,[5]Ahmad Samed Al-Adwan

[1]Senior Lecturer (English) Department of computer Science Lahore Garrison University

Uzmashaheen@lgu.edu.pk

[2]marriam49@gmail.com

Al Munsiyah, Riyadh, Saudi Arabia The University of Lahore

[3]Department of English Language and Literature

ammara.afzaal@ell.uol.edu.pk

[4]AI R&D Researcher Computing and Information Sciences Egypt University of Informatics

Cairo, Egypt Omarshafiy15@gmail.com

23-201356@students.eui.edu.eg

[5]Business school, Department of Business Technology, Al-Ahliyya Amman University, Amman, Jordan

a.adwan@ammanu.edu.jo

orcid.org/0000-0001-5688-1503

HOW TO CITE:

Uzma Shaheen, Mariam Saleem,Ammara Afzaal,Omar M. Shafiy,Ahmad Samed Al-Adwan(2026). An Intelligent Decision Support Model Using Hybrid K-Means PSO-SVM for MSME Development Priority Classification

International Journal of Special Education, 41 (21s) 612 - 638

ABSTRACT

Generative AI (GAI) has come into English classrooms before the evidence base to regulate it, and nearly all the evidence surrounding GAI treats the classroom as a neutral receptacle. This study examined whether the benefits of AI-mediated instruction are related to the architecture of the space in which the instruction is delivered, and what the instruction does to learners' cultural identity. The study employed a 16-week quasi-experimental design with convergent qualitative strand with four universities of public sector in Punjab, Pakistan. Twenty-four intact undergraduate sections (N = 597 students) were assigned to AI mediated or conventional instruction, in either an Adaptive Learning Studio (reconfigurable furniture, distributed display surfaces, high illuminance, low reverberation) or a Conventional Lecture Room. Outcomes were literary interpretive competence (blind double-rated analytic essays), four-dimensional engagement, digital/AI literacy and three facets of cultural identity negotiation; indoor environmental quality was measured objectively in every room. AI-mediated instruction produced large gains in digital/AI literacy ($\eta^2p = .241$, $g = 1.10$ [0.93, 1.27]), engagement ($\eta^2p = .182$) and interpretive competence ($\eta^2p = .216$), but the instruction × environment interaction was significant for attainment, $F(1, 592) = 37.22$, $p < .001$, $\eta^2p = .059$:

the AI effect was more than twice as large in the studio (g = 0.95 [0.71, 1.20]) as in the conventional room (g = 0.40 [0.17, 0.63]). The indirect effect of serial mediation (AI → digital literacy → engagement → attainment) was 0.88 points [0.67, 1.14] and the index of moderated mediation was 0.52 [0.10, 0.99]. While there was a slight increase in perceived cultural dilution (g = 0.21) and a slight decrease in heritage cultural affirmation (g = −0.27), there was a medium increase in awareness of AI and Global Englishes (g = 0.45), with both of the latter results mitigated by a stronger ethical AI-literacy (r = −.30, p < .001). The findings suggest that AI integration and learning space design should be considered as a single intervention and that the mediational and cultural pathways need to be experimented and demonstrate a correlational relationship.

Keywords: AI in education, English language teaching, digital literacy, cultural identity, student engagement, learning environment design, indoor environmental quality, Global Englishes.

1. Introduction

The rapid and uneven introduction of generative AI to the field of English studies

Generative AI, as embodied in large language model (LLM) interfaces, is now a ubiquitous part of English classroom learning for undergraduates, within three years of the public launch of LLM interfaces. Over the last two years, hundreds of empirical studies have been documented through systematic reviews, which all agree that AI tools have a positive, reliable impact on language learning on writing fluency, foreign language anxiety and reported motivation (Alhusaiyan, 2025; Crompton et al., 2024; Wang et al., 2025). The results of meta-analytic syntheses indicate positive but highly heterogeneous results for achievement (Liu et al., 2025; Wu & Li, 2024) and higher-order thinking (Wang & Fan, 2025; Xia et al., 2025). The interesting part is the heterogeneity. The reported effect sizes of AI in language learning vary from small to large and the moderators studied so far are tool type, outcome measure, and length of intervention, which cannot explain the variation (Liu et al., 2025; Lo et al., 2024). There is something systematic, missing from the models.

The physical setting is one of the candidates that could have been omitted. Historically, educational technology research has taken the room as a neutral container, and the intervention is the software; the room is where the software happens. However, there is an established and largely independent body of literature which shows that the built environment has measurable effects on attention, comfort and short term academic success. Brink et al. (2021) meta analysed 21 high-quality studies of higher education and found a positive correlation between IEQ and quality of learning and a positive correlation between IEQ and short-term academic performance; in a between-groups field experiment, reducing the reverberation time and increasing horizontal illuminance to Dutch class-A standards resulted in better perception of IEQ and better problem solving abilities among students (Brink et al., 2023). Research on parallel work on layout, not ambient quality, reveals that circular and reconfigurable layouts can be more "immediate" in that the flow of feedback exchange is more immediate, and that active-learning classrooms can facilitate collaborative configurations not found in fixed-row classrooms (Parsons, 2024; Bingen et al., 2023). Classroom environment has recently emerged as a missing antecedent in learning engagement in the context of English as a foreign language (EFL) in specific.For English as a foreign language (EFL), classroom environment has recently started to be discussed as a missing antecedent of learning engagement (Ye, 2024). To the best of our knowledge, however, no experimental cross-over between the two literatures exists.

1.2 Four constructs and why they are included in one model.

This study combines four constructs in one design as they have been demonstrated to be

significant on their own and as their relationship is theoretically compelling.

Digital and AI literacy. AI literacy is beyond tool mastery. Ng et al. (2024) operationalized it at the affective, behavioural, cognitive and ethical (ABCE) levels, and found that ethical reasoning about AI is independently measurable from its use. Other research has shown that digital access and computational thinking are determinants of AI literacy, in addition to exposure (Celik, 2023) and that measured AI competence at entry is predictive of intended and actual use of the tools (Delcker et al., 2024). The skill that's being measured is how to determine when to reject an AI suggestion when they're not intended for use in writing-intensive fields (Cardon et al., 2023).

Student engagement. While engagement has been the most consistent proximal benefit reported for the use of AI (Deep et al., 2025; Derakhshan, 2025; Youssef et al., 2024), in a systematic review, Lo et al. (2024) address the construct in terms of undifferentiated enthusiasm. In this context treating engagement multidimensionally (behavioural and emotional, cognitive, agentic) makes sense, because the physical environment is plausibly more of a catalyst in some (agentic, behavioural) than in others (cognitive) dimensions.

Cultural identity. LLMs are trained on corpora containing over-represented standardised native-speaker Englishes. This was experimentally shown by Lee et al. (2025) who found that the Korean English features—such as borrowings, localised syntax, culturally embedded pragmatic strategies—that the unmodified ChatGPT model failed to generate were also features of Korean English, showing that the model lacks support for these features. AI's monolingual output has been criticized for perpetuating monolingual ideologies and marginalizing minority language variants (Dai et al., 2025; Jenks, 2025; Jeon et al., 2025). This is a core issue, not a peripheral one, for multilingual learners in postcolonial settings, where the status of the English language is already contested. It is also not well tested quantitatively.

Architectural learning environment. We consider the room to be an adjustable variable in two aspects: measurable ambient quality (illuminance, CO2 concentration, reverberation time, background noise); spatial affordance (furniture reconfigurability, display surface density, circulation). Both are evidentially supported (Brink et al., 2021, 2022; Parsons, 2024) and both are likely required for the collaborative critique activities on which AI pedagogy relies (Brink et al., 2021, 2022; Parsons, 2024).

1.3 The gap and the argument of this paper

This study is thus a specific study and not a generic one. The truth is that there is plenty of research on AI in education in general, and in English specifically, but that the research is architecturally blind. It is not clear that the pedagogical importance of AI in literature teaching relies on the ability of students to gather in small groups, observe each other and show their work in a small group setting, and if it does, then that would require a room where small groups can gather, see each other, and show their work. Then, room design is a moderator and should be modelled rather than being a covariate to be controlled. Another area of concern in relation to culture is the critical literature arguing that AI-based English teaching may lead to a uniform voice of the learners (Jeon et al., 2025; Lee et al., 2025), which is based almost entirely on the study of outputs from the models and not on the measurement of what happens to the learners.

1.4 Theoretical framework

The model used in this test (Figure 1) has the following three positions. In a sociocultural point of view, AI is seen as a mediating artefact, which does not directly influence the learner but transforms the activity through which learning takes place, and thus affects it depending on other components of the activity system—including its material setting. In terms of self-determination, engagement is fostered when the learners feel competent, autonomous and connected; AI can provide competence feedback; autonomy and relatedness are more spatial goods. The pedagogical object is not conformity to a standard variety, but the capacity to move critically among varieties, a perspective from a Global Englishes standpoint (Lee et al., 2025), which shifts the role of the cultural outcomes to another part of the same model as the cognitive ones, rather than to an ethical appendix.

We, therefore, define AI-mediated instruction as a distal cause, digital/AI literacy and engagement as serial mediators of attainment, cultural identity negotiation as a parallel outcome, and the architectural environment as a main effect as well

as a moderator of the engagement–attainment pathway.

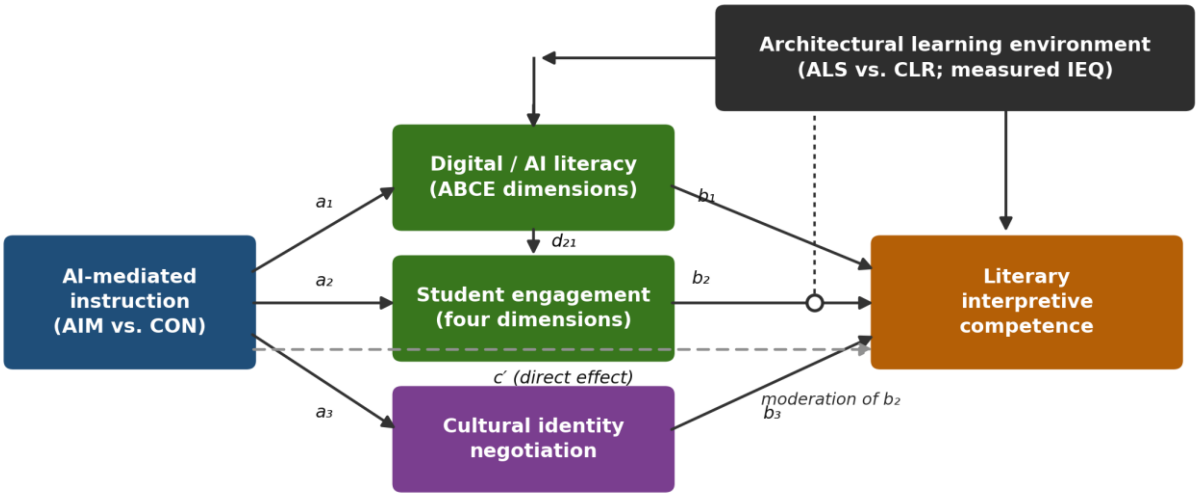

Figure 1. Hypothesised model relating AI-mediated instruction, digital/AI literacy, student engagement, cultural identity negotiation and literary interpretive competence, with the architectural learning environment as moderator.

Note. ALS = Adaptive Learning Studio; CLR = Conventional Lecture Room; IEQ = indoor environmental quality; ABCE = affective, behavioural, cognitive, ethical. The dashed line denotes the direct effect; the open circle denotes moderation of the engagement→attainment path.

1.5 Aims and hypotheses

Six a priori hypotheses were tested.

- H1. Compared to conventional instruction, AI mediated instruction results in greater post-intervention LIC after accounting for pre-test performance.
- H2. AI-powered teaching has a positive impact on all four dimensions of engagement.
- H3. The instructional effects of AI result in increased digital/AI literacy, especially with the cognitive and ethical aspects.
- H4. The use of AI-based teaching is correlated with increased awareness of Global Englishes, as well as decreased heritage cultural affirmation and increased perceived cultural dilution.
- H5. The Conventional Lecture Room is not as effective as the Adaptive Learning Studio, and the effect of AI-mediated instruction (instruction × environment interaction) is amplified by the Adaptive Learning Studio.
- H6. The impact of AI-mediated teaching on achievement is mediated through digital/AI literacy and engagement which is in turn mediated by the learning environment.

H1–H3 and H5 are causal claims that have random assignment at the section level. Measured variables, as opposed to manipulated variables are used in H4 and H6 and these are treated throughout as associational, a distinction made on purpose in the Results, and discussed again in the Discussion.

2. Materials and Methods

2.1 Design

The design adopted was a convergent mixed-methods design that involved a 2 (instruction: AI-mediated [AIM] vs. conventional [CON]) × 2 (environment: Adaptive Learning Studio [ALS] vs. Conventional Lecture Room [CLR]) cluster-randomised quasi-experiment and an embedded qualitative strand. Randomisation was done at class section level as the manipulation of the environment is a group level assignment. Both strands of the qualitative and quantitative were accorded equal importance and were merged at the level of interpretation. The intervention was for one semester of 16 weeks.

2.2 Setting and participants

The participants were four public sector universities, in Punjab, Pakistan. All four provide a 4-year BS in English (Language and Literature), and share a similar Higher Education Commission aligned curriculum that allowed for common instruments and a common assessment rubric across sites. Thirty-two sections were tested, 24 of which met the inclusion criteria (20–30 students; both courses taught by the same instructor; both room types available in the same learning schedule; random selection of 6 sections in each of 4 cells, stratified by university to ensure each site contributed a section to each cell).

There were 641 enrolled students, of which 44 were excluded (17 withdrew, 19 had below 75% attendance and 8 did not take the post-test) leaving an analysed sample of N = 597 in k = 24 sections (Figure 2). Attrition did not differ across conditions, $\chi^2(3) = 2.41$, $p = .492$. Participants had a mean age of 20.63 years (SD = 1.46); 58.8% were women. The characteristics of the samples are presented in Table 1. Using G*Power conventions, the design was sensitive to detect a between condition standardised difference of ~ d = 0.45 with 80% power for k = 24 clusters, an average cluster size of 25 and an intraclass correlation of .23 at $\alpha = .05$.

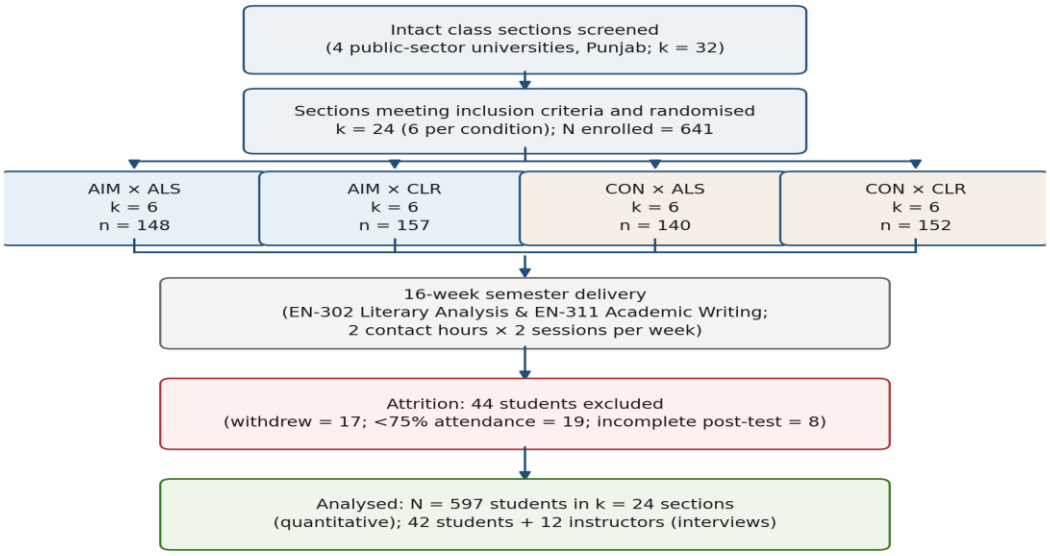

Figure 2. Participant flow from section screening to analysed sample.

Table 1 *Characteristics of the analysed sample (N = 597)*

Characteristic	n	%
Age (years), M (SD)	20.63 (1.46)	—
Gender		
Women	351	58.8
Men	246	41.2
Year of study		
Second year	177	29.6
Third year	230	38.5
Fourth year	190	31.8
First language		

Characteristic	n	%
Punjabi	253	42.4
Urdu	182	30.5
Saraiki	113	18.9
Pashto or other	49	8.2
Home locality		
Urban	283	47.4
Semi-urban	187	31.3
Rural	127	21.3
Device access		
Smartphone only	262	43.9
Smartphone and laptop	335	56.1
Prior generative AI use		
None or rare	162	27.1
Occasional	295	49.4
Frequent	140	23.5

2.3 The instructional manipulation

There was no difference between the two conditions with regard to content or main texts, assessment weightings or total contact time (4 hours/week). The manipulation involved in the generation, interrogation and revision of text and not what was read. The two conditions and two room types to be detailed for replication are defined in Table 2.

In the AIM condition, every two-hour session consisted of a 35-minute "interrogate-the-machine" cycle that involved student pairs (a) asking an AI to read a passage, using a structured prompt frame; (b) marking up the AI's output against that of the primary text for interpretive overreach, unsupported generalisation and cultural flattening; (c) producing a counter-reading in their own voice; and (d) posting both to a shared display surface for cross-group critique. All instructors followed a common facilitation approach, and it was explicitly stated they are not allowed to accept AI answers as the definitive answer. In the CON condition, the students were given 35 minutes in a similar close reading and critique cycle but were provided with a set of critical excerpts written by the teacher instead of AI-generated content, while maintaining the collaborative nature and leaving out only the AI. This is an active-control design, which separates out AI mediation and not blurred by the collaborative pedagogies.

Fidelity was assessed by fortnightly observation of 25% of sessions by two trained observers on a 14 item protocol checklist. Mean fidelity was 0.89 (SD = 0.07) in AIM sections and 0.91 (SD = 0.06) in CON sections; inter-observer agreement was $\varkappa$ = 0.87. Teachers were given a standard 12 hours training and did not teach in more than one condition.**Table 2**

Specification of the instructional and architectural conditions

Feature	AI-mediated (AIM)	Conventional (CON)
Text-generation source	LLM chat interface with structured prompt frame	Instructor-prepared critical excerpts

Feature	AI-mediated (AIM)	Conventional (CON)
Core weekly cycle	Elicit → annotate → counter-read → display → critique	Read → annotate → counter-read → display → critique
Feedback on drafts	AI first-pass plus instructor and peer review	Peer review plus instructor review
Explicit AI-literacy teaching	90 min in Weeks 1–2 plus embedded prompts on bias, hallucination, attribution	None
Attribution requirement	Mandatory AI-use log appended to every submission	Not applicable
Contact hours per week	4	4
Assessment weighting	Identical	Identical
Architectural feature	**Adaptive Learning Studio (ALS)**	**Conventional Lecture Room (CLR)**
Seating	Movable castor chairs; 6 reconfigurable pods	Fixed rows of bench desks
Writable / display surfaces	4 wall-mounted whiteboards plus 2 projection walls	1 front whiteboard plus 1 projector
Sightlines	Multi-focal; no designated front	Uni-focal toward instructor
Target horizontal illuminance	≥ 500 lx at desk plane	Uncontrolled (as built)
Acoustic treatment	Ceiling baffles; target $RT_{60} \leq 0.6$ s	Untreated
Ventilation	Mechanical, demand-controlled	Natural (openable windows)
Floor area per student	≈ 2.4 m²	≈ 1.3 m²

2.4 Measures

Competence in literary interpretation (LIC). At the beginning of the program (Week 1) and at the end (Week 16), participants completed a 900–1,100 word analytic essay on an unseen poem (poems counterbalanced across pre- and post-test). Essays were marked using a 40-point analytic rubric consisting of five 8-point dimensions: text evidence, interpretive argument, contextual and cultural reading, structural control, linguistic precision. All scripts were independently rated by two rateers, blind to condition and time point, with the intraclass correlation, two-way random, absolute-agreement, average-measures ICC(2,k) = 0.91 [0.88, 0.93]. More than 4 points discrepancy (4.1% of scripts) were judged by a third rater.

Digital/AI literacy. An adaptation of the AI Literacy Questionnaire (Ng et al., 2024) was used, comprising 24 items with the same affective, behavioural, cognitive and ethical structure, but with items reworded for use in an English-studies context (AILQ-EL). Items were rated 1–5. The two items were deleted and replaced with items on evaluating AI-generated literary interpretation based on the construct definition, not wording.

Student engagement. The SES-4D was a 21 item, 1–5 scale, measuring behavioural, emotional, cognitive and agentic engagement with the literature seminar, which were adapted from literature.

Cultural identity negotiation. A 16-item instrument (CIN-E) was created for this study from the literature of the Global Englishes (Jenks, 2025; Lee et al., 2025) and expert review (six applied linguists) and cognitive interviews with 14 students. It consists of heritage cultural affirmation (6 items; e.g., confidence in using Punjabi, Saraiki or Urdu literary reference in English writing), Global Englishes awareness (5 items) and perceived cultural dilution (5 items; e.g., the feeling that one's written voice is

becoming indistinguishable from a machine's voice).

Perceived spatial affordance. Students' perceptions of the room's ability to reconfiguration, visibility of others' work, acoustic and physical comfort were captured with a 7-item scale.

Objective IEQ. Calibrated instruments were used to record horizontal illuminance (lux, desk plane), CO2 concentration (ppm), reverberation time (RT60, s), and A-weighted background noise (dB), within all of the teaching rooms at 10-minute intervals for 6 randomly selected sessions per section, while occupied floor area per student was measured once. A standardised IEQ composite index was calculated as the mean of z-scored illuminance and area per student, minus z-scored CO2 and RT60 and noise, with a higher score indicating better conditions.

Each of the scales was administered in English, the medium of instruction, at Week 1 and Week 16. The properties of reliability and distributional are reported in Table 3

.**Table 3** *Internal consistency and distributional properties of the measures at post-test*

Scale (number of items)	k	α	ω	M	SD	Skew	Kurtosis
AILQ-EL: Affective (6 items)	6	0.89	0.89	3.61	0.69	-0.03	-0.60
AILQ-EL: Behavioural (6 items)	6	0.88	0.88	3.47	0.70	-0.05	-0.26
AILQ-EL: Cognitive (7 items)	7	0.90	0.90	3.44	0.71	-0.07	-0.32
AILQ-EL: Ethical (5 items)	5	0.86	0.86	3.38	0.75	-0.11	-0.17
SES-4D: Behavioural (5 items)	5	0.88	0.88	3.66	0.67	-0.05	-0.49
SES-4D: Emotional (5 items)	5	0.89	0.89	3.53	0.73	-0.08	-0.33
SES-4D: Cognitive (6 items)	6	0.89	0.89	3.55	0.70	-0.17	-0.29
SES-4D: Agentic (5 items)	5	0.85	0.85	3.43	0.76	-0.10	-0.39
CIN-E: Heritage affirmation (6 items)	6	0.88	0.88	3.65	0.67	-0.11	-0.13
CIN-E: Global Englishes awareness (5 items)	5	0.85	0.85	3.36	0.70	-0.06	-0.25
CIN-E: Perceived cultural dilution (5 items)	5	0.86	0.86	3.03	0.77	-0.11	-0.11
Perceived spatial affordance (7 items)	7	0.88	0.88	3.05	1.06	-0.00	-1.40

2.5 Qualitative strand

Forty-two students (stratified by condition, gender and attainment tercile) and 12 instructors took part in semi-structured interviews of 35–50 minutes in Weeks 14–16. The timeline explored the use of or lack of use of AI, the nature of voice and authorship, and how the space influenced the way that sessions went. The interviews were held in the interviewee's preferred language (English, Urdu or Punjabi), audio taped, transcribed and translated (where needed), with back-translations for checks. Two coders independently analyzed the data using reflexive thematic analysis, with an inter-coder agreement at the final codebook of $\varkappa$ = 0.84. Disagreements were sorted out in discussion and a traceability of coding decisions was kept.

2.6 Statistical analysis

All analyses were performed in Python 3.12 with the packages statsmodels 0.14 and SciPy 1.17. The analysis plan was predetermined prior to the analysis of the post-test data. The following assumptions were tested before each test: normality of residuals was checked by observing quantile–quantile plots and skew/kurtosis indices, homogeneity of variance was checked with Levene's test, and homogeneity of

regression slopes was checked by testing the covariate × factor interaction.

2 x 2 analysis of variance was used to test the primary hypotheses with pre-test score as a covariate for the attainment outcome. The effect sizes for the omnibus tests are reported as partial eta squared and for the pairwise comparisons as Hedges' g with 95% confidence intervals. All of the attainment models were estimated as linear mixed models with a random section intercept, since students were nested within sections, and the unconditional intraclass correlation was calculated from the null model. Ordinary least squares path models and percentile bootstrap confidence intervals (5,000 resamples) were estimated for serial mediation and for moderated mediation. The p value is reported exactly to three decimal places, or as $p < .001$, if it is smaller, and the pattern of confidence intervals is interpreted, not only significance thresholds. The six pre-specified hypotheses were stated without making any correction for multiple testing; the exploratory subgroup analyses described in Section 3.9 are identified as such, and should be interpreted accordingly.

2.7 Ethics

The protocol was accepted by the institutional review board of the coordinating university and by the research committees of the three partner sites. Written informed consent was obtained from all participants and participants in the CON condition were given a four session AI-literacy workshop following the post-test, while all transcripts of the AI were anonymised prior to analysis. No course credit was awarded for participation and no academic penalty existed for not participating.

3. Results

A series of preliminary analyses and assumption checks were conducted.A list of preliminary analyses and assumption checks were performed.

Groups were equivalent at baseline due to randomisation. Pre-test literary interpretive competence did not differ by instruction, $F(1, 593) = 1.81$, $p = 0.179$, by environment, $F(1, 593) = 1.26$, $p = 0.262$, or by their interaction, $F(1, 593) = 0.44$, $p = 0.505$. There were also no differences in age, gender composition, first language, home locality, or device access or previous AI use across the four cells (all p values $> .18$).

Assumptions of the covariance analysis were met. The Levene's test for homogeneous post-test variance across the four cells was not rejected ($W(3, 593) = 1.93$, $p = 0.124$), and the three-way covariate × instruction × environment term was not significant ($p = 0.437$), indicating homogeneity of regression slopes. Absolute skews and kurtosis were near normal (≤ 0.31 and ≤ 0.42, respectively) for residuals. The variance inflation factors of the largest regression models ranged from 1.01 to 5.94; two of these are higher than normal, with the environment dummy (5.94) and the perceived spatial affordance (5.65), but these factors are well below the commonly cited threshold of 10, and both are related to each other and should not be interpreted separately.

Harman's single factor test was conducted for common-method variance with the first unrotated factor explaining 28.4% variance which is below 50% criteria. The scalar level model invariance ($\Delta CFI \leq .004$ and $\Delta RMSEA \leq .001$ each step) supports the between-cell mean comparisons conducted below.

3.2 Descriptive statistics

The means and standard deviations for each outcome are presented in Table 4. Two aspects of the descriptive pattern look forward to the inferential outcomes. First, the attainment measure × ALS cell ($M = 28.01$) is different from the other three cells which are 23.83, 23.72, and 22.27, and are within 1.6 points of each other. This is the signature of an interaction not two additive main effects. Second, the cultural identity outcomes are different: In the case of Global Englishes awareness, it is increased with the AI mediation while for heritage affirmation it is reduced, therefore, the "cultural" impact of the AI is not one dimensional.

Table 4 *Cell means and standard deviations for all outcome measures at post-test*

Measure	AIM × ALS (n = 148)	AIM × CLR (n = 157)	CON × ALS (n = 140)	CON × CLR (n = 152)	Total (N = 597)
Literary interpretive competence, pre-test (0–40)	21.53 (4.17)	22.10 (4.01)	22.18 (3.70)	22.32 (3.83)	22.03 (3.94)
Literary interpretive competence, post-test (0–40)	28.01 (4.67)	23.83 (4.17)	23.72 (4.30)	22.27 (3.60)	24.44 (4.70)
Engagement, composite (1–5)	4.03 (0.38)	3.45 (0.43)	3.49 (0.43)	3.22 (0.37)	3.55 (0.50)
Behavioural engagement	4.13 (0.57)	3.55 (0.63)	3.55 (0.62)	3.42 (0.62)	3.66 (0.67)
Emotional engagement	4.06 (0.61)	3.42 (0.72)	3.47 (0.64)	3.19 (0.63)	3.53 (0.73)
Cognitive engagement	3.95 (0.63)	3.54 (0.63)	3.51 (0.69)	3.21 (0.64)	3.55 (0.70)
Agentic engagement	4.00 (0.58)	3.26 (0.68)	3.42 (0.78)	3.06 (0.68)	3.43 (0.76)
Digital/AI literacy, composite (1–5)	3.82 (0.38)	3.57 (0.40)	3.28 (0.42)	3.22 (0.38)	3.47 (0.46)
Affective	4.02 (0.62)	3.70 (0.62)	3.31 (0.66)	3.40 (0.65)	3.61 (0.69)
Behavioural	3.80 (0.71)	3.51 (0.64)	3.34 (0.67)	3.21 (0.63)	3.47 (0.70)
Cognitive	3.73 (0.63)	3.60 (0.75)	3.18 (0.66)	3.24 (0.65)	3.44 (0.71)
Ethical	3.72 (0.70)	3.48 (0.72)	3.30 (0.71)	3.01 (0.68)	3.38 (0.75)
Heritage cultural affirmation (1–5)	3.62 (0.65)	3.50 (0.67)	3.76 (0.71)	3.72 (0.63)	3.65 (0.67)
Global Englishes awareness (1–5)	3.52 (0.74)	3.51 (0.68)	3.24 (0.61)	3.16 (0.68)	3.36 (0.70)
Perceived cultural dilution (1–5)	3.11 (0.74)	3.10 (0.79)	2.87 (0.80)	3.02 (0.72)	3.03 (0.77)

3.3 Verification of the architectural manipulation

The environmental manipulations were not assumed but actually verified. Table 5 demonstrates that the two room types were significantly different in all parameters measured. Mean horizontal illuminance at the desk plane was 626 lx in the studios against 358 lx in the conventional rooms, $t(22) = 13.9$, $p < .001$; mean CO2 concentration was 875 ppm against 1350 ppm, $t(22) = -10.39$, $p < .001$; and mean reverberation time was 0.52 s against 0.89 s, $t(22) = -11.87$, $p < .001$. Only the studios achieved illuminance values within the target value set in the intervention condition, and only the studios were able to maintain CO2 below 1,000 ppm, which is a widely used proxy for adequate ventilation. In 71% of the recorded periods, the conventional rooms surpassed it. These contrasts are very large by convention, which is to be expected: they are man-made differences between two building types, not the natural variation, which should be interpreted as a manipulation check, not as findings.

Table 5

Objectively measured indoor environmental quality by room type (section level, k = 24)

Parameter	Target	ALS M	SD	CLR M	SD	t(22)	p	Hedges' g
Horizontal illuminance (lx)	≥ 500	625.5	34.37	358.4	57.00	13.90	< .001	5.48

Parameter	Target	ALS M	SD	CLR M	SD	t(22)	p	Hedges' g
CO_2 concentration (ppm)	< 1,000	875.0	58.45	1350.0	147.22	-10.39	< .001	-4.09
Reverberation time RT_{60} (s)	≤ 0.60	0.52	0.08	0.89	0.07	-11.87	< .001	-4.68
Background noise, L_{aea} (dB)	≤ 45	41.4	2.89	52.2	3.56	-8.19	< .001	-3.23
Floor area per student (m^2)	≥ 2.0	2.39	0.21	1.29	0.12	15.93	< .001	6.28

3.4 H1 and H5: Literary interpretive competence

The 2 × 2 analysis of covariance on post-test attainment, with pre-test score as covariate, yielded significant main effects of instruction, $F(1, 592) = 163.41$, $p < .001$, $\eta^2p = 0.216$, and of environment, $F(1, 592) = 154.51$, $p < .001$, $\eta^2p = 0.207$, together with a significant instruction × environment interaction, $F(1, 592) = 37.22$, $p < .001$, $\eta^2p = 0.059$. The model explained 58.2% of the variance in the post-test scores. Both main effects were supported, and the interaction is also substantively important and qualifies the main effects. Qualification is clarified with simple-effects analysis (Table 7, Figure 3a). In the Adaptive Learning Studio, AI-mediated instruction produced a mean advantage of 4.3 points on the 40-point scale, 95% CI [3.26, 5.33], $t(286) = 8.13$, $p < .001$, $g = 0.954$ [0.71, 1.198]. In the Conventional Lecture Room the same instructional manipulation produced an advantage of only 1.56 points [0.69, 2.43], $t(307) = 3.53$, $p < .001$, $g = 0.4$ [0.174, 0.625]. The two simple effects have no overlap in their confidence intervals. The pedagogy assumes that the room allows for collaborative activities, which when expressed as a ratio, resulted in a 2.8 times larger instruction effect. The difference that is most relevant to practice is between the AI-mediated studio cell (M = 28.01) and the conventional instruction in a conventional room cell (M = 22.27) which is 5.74 points. Specifically, the AI mediated conventional room cell (M = 23.83) and the conventional instruction studio cell (M = 23.72) did not significantly differ, $t(295) = 0.23$, $p = .820$. With these conditions, if AI were added to an inappropriate room, the results were about the same as if the room were improved without AI — and the results were even better if both AI and room improvement were used

.

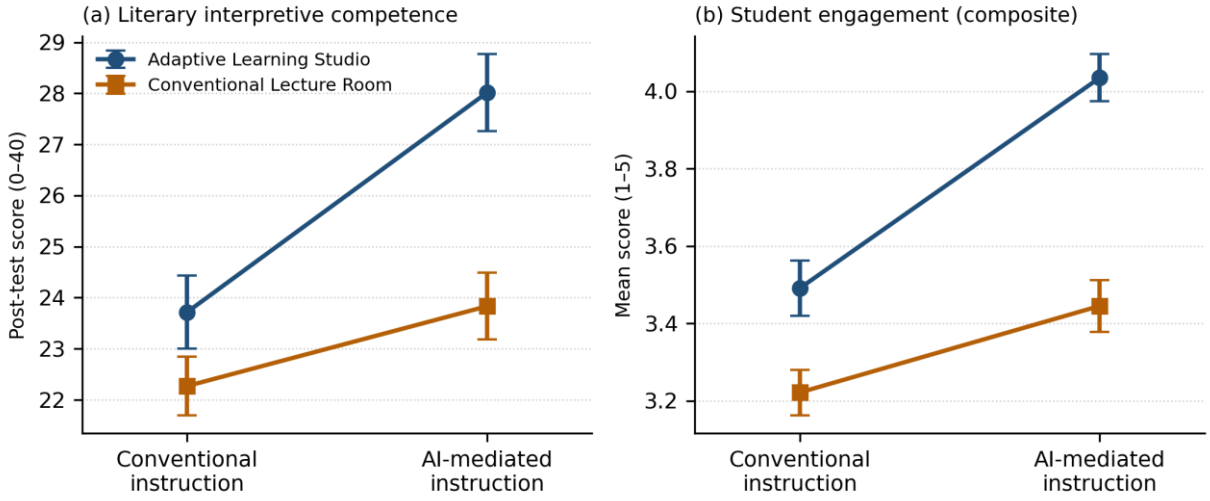

Figure 3. Instruction × environment interaction for (a) literary interpretive competence and (b) composite student engagement.

Table 6 *Summary of 2 × 2 factorial analyses for all outcomes*

Outcome	F(AI)	η^2p	F(Env)	η^2p	F(AI×Env)	η^2p	R^2
Literary interpretive competence	163.41	0.216	154.51	0.207	37.22	0.059	0.582
Engagement (composite)	131.60	0.182	172.56	0.225	23.56	0.038	0.357
Behavioural	48.66	0.076	52.61	0.081	19.46	0.032	0.170
Emotional	55.86	0.086	74.95	0.112	11.18	0.019	0.194
Cognitive	52.60	0.081	45.56	0.071	0.97	0.002	0.144
Agentic	46.22	0.072	96.52	0.140	11.35	0.019	0.207
Digital/AI literacy (composite)	188.10	0.241	23.58	0.038	7.77	0.013	0.271
Affective	89.20	0.131	5.33	0.009	15.61	0.026	0.157
Behavioural	49.74	0.077	15.06	0.025	2.24	0.004	0.102
Cognitive	67.47	0.102	0.46	0.001	3.01	0.005	0.107
Ethical	60.50	0.093	20.53	0.033	0.15	0.000	0.121
Heritage cultural affirmation	10.96	0.018	2.06	0.003	0.56	0.001	0.022
Global Englishes awareness	30.67	0.049	0.68	0.001	0.44	0.001	0.051
Perceived cultural dilution	6.40	0.011	1.32	0.002	1.67	0.003	0.016

Table 7 *Simple effects of AI-mediated instruction within each architectural environment*

Contrast (AIM − CON)	M diff	t	p	Hedges' g	95% CI
Literary interpretive competence — Adaptive Learning Studio	4.30	8.13	< .001	0.95	[0.71, 1.20]
Literary interpretive competence — Conventional Lecture Room	1.56	3.53	< .001	0.40	[0.17, 0.62]
Engagement (composite) — Adaptive Learning Studio	0.54	11.34	< .001	1.34	[1.08, 1.59]
Engagement (composite) — Conventional Lecture Room	0.22	4.93	< .001	0.56	[0.33, 0.79]

3.5 H2: Student engagement

AI-mediated instruction raised composite engagement substantially, $F(1, 593) = 131.6$, $p < .001$, $\eta^2p = 0.182$, $g = 0.82$ [0.653, 0.987]. The environmental main effect on engagement was, if anything, larger, $F(1, 593) = 172.56$, $p < .001$, $\eta^2p = 0.225$, $g = 0.963$ [0.794, 1.133], and the interaction was again significant, $F(1, 593) = 23.56$, $p < .001$, $\eta^2p = 0.038$ (Figure 3b). H2 was supported. The subscale pattern is more informative than the composite and is best evidence from the data set in support of a mechanism, as opposed to a general positivity effect (Figure 4, Table 6). The instructional main effect was very comparable across the four dimensions (η2p from 0.072 to 0.086), indicating that it was a general effect of AI on engagement. The environmental main effect was not consistent; it was largest for agentic engagement ($\eta^2p = 0.14$), and emotional engagement ($\eta^2p = 0.112$) was smallest. The interaction was significant for behavioural, emotional and agentic engagement, but not for cognitive engagement, $F(1, 593) = 0.97$, $p = 0.325$. This disconnection is logically consistent. Agentic engagement – students starting the task, redirecting it, and claiming it – depends on the physical affordances of being able to turn towards a peer, move a chair, and claim a display surface. The extent of processing to a text is a more private operation that is less inhibited by a fixed-row room, called cognitive engagement. This distinction is reinforced by the fact that the effect in the room follows this distinction rather than affecting all four dimensions of the room equally, making such an effect unlikely to be a novelty or demand-characteristic effect.

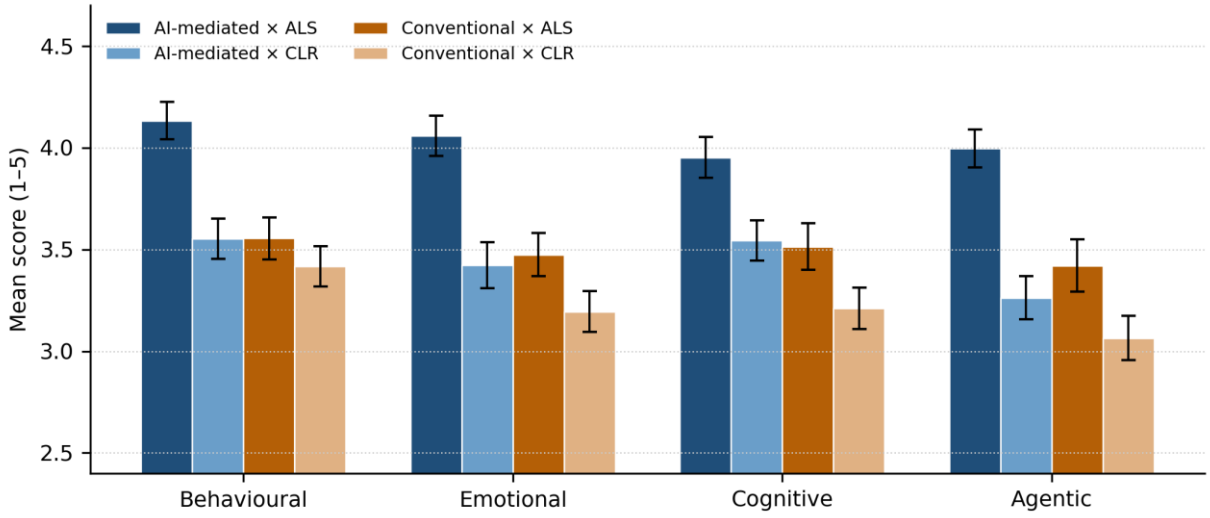

Figure 4. Engagement subscale means by instruction and environment.

3.6 H3: Digital and AI literacy

The largest instructional effect in the study was on digital/AI literacy, $F(1, 593) = 188.1$, $p < .001$, $\eta^2p = 0.241$, $g = 1.097$ [0.925, 1.269]. This is the result most directly linked to the manipulation and most intuitively expected: students who had a semester of experience in eliciting, annotating and challenging machine output said they could do more of it. H3 foretold the increase in the cognitive and ethical aspects would be the greatest, which has been partially true. The affective dimension showed the largest effect ($\eta^2p = 0.131$, $g = 0.761$), followed by cognitive ($\eta^2p = 0.102$), ethical ($\eta^2p = 0.093$) and behavioural ($\eta^2p = 0.077$) dimensions (Figure 5a). Thus, H3 is recorded as partially supported: the direction was as predicted for all dimensions, but the predicted ordering was not achieved. The environment added a small independent effect to literacy ($\eta^2p = 0.038$) with a focus on the behavioural ($\eta^2p = 0.025$) and ethical ($\eta^2p = 0.033$) dimensions, but not on the cognitive dimension ($\eta^2p = 0.001$, $p = 0.497$). The interview data in Section 3.10 suggests a plausible reading that visible peer work socializes ethics around attribution that private screening doesn't. This reading is NOT an established mechanism, but a hypothesis formed based on the data. 3.7 H4: Cultural identity negotiation The cultural outcomes were most ambiguous and, indeed, most significant, as we will argue. Figure 5b shows that H4 predicted a three-part pattern and

achieved that, but with effect sizes considerably smaller than those for cognition and engagement. Global Englishes awareness was higher under AI mediation, $F(1, 593) = 30.67$, $p < .001$, $\eta^2p = 0.049$, $g = 0.454$ [0.291, 0.616]. Heritage cultural affirmation was lower, $F(1, 593) = 10.96$, $p = 0.001$, $\eta^2p = 0.018$, $g = -0.27$ [-0.431, -0.109]. Perceived cultural dilution was higher, $F(1, 593) = 6.4$, $p = 0.012$, $\eta^2p = 0.011$, $g = 0.206$ [0.045, 0.367]. None of these outcomes had any significant environmental main effect or interaction (all $F < 2.10$, all $p > .15$), which is itself a finding: room design enhanced the ways that students worked, but they did not find it measurably helpful in terms of maintaining their sense of cultural voice. There are two qualifications which restrict the extent of this. First, both the affirmation and dilution effects are small (both $|g| < 0.30$) and the lower bound of the dilution confidence interval [0.045, 0.367] is close to 0, a change that occurs over a large sample, not an identity change over one semester. Second, there was no negative relationship between awareness and affirmation in the overall sample ($r = 0.191$, $p < .001$), which means that the pattern of results is not strictly a trade-off between affiliation to World Englishes and to one's own. The most actionable result in this section is an unplanned analysis that limited to the AI-mediated group. Ethical AI literacy was negatively associated with perceived cultural dilution, $r = -0.296$, 95% CI [−0.40, −0.19], $p < .001$, $n = 305$, whereas behavioural (usage-oriented) AI literacy was not ($r = .09$, $p = .121$). Split by composite literacy tercile, mean dilution fell from 3.33 (SD = 0.78) in the lowest tercile to 3.03 (SD = 0.77) in the highest, $F(2, 302) = 2.77$, $p = 0.064$. A Tercile contrast is not significant at $\alpha = .05$ and should not be reported as such and is not the more reliable result; the continuous correlation with the ethical subscale is the more reliable. Combined, these analyses argue — without proving — that it is necessary and right that it is AI literacy, rather than knowledge of AI tools, that goes hand-in-hand with a safe voice.

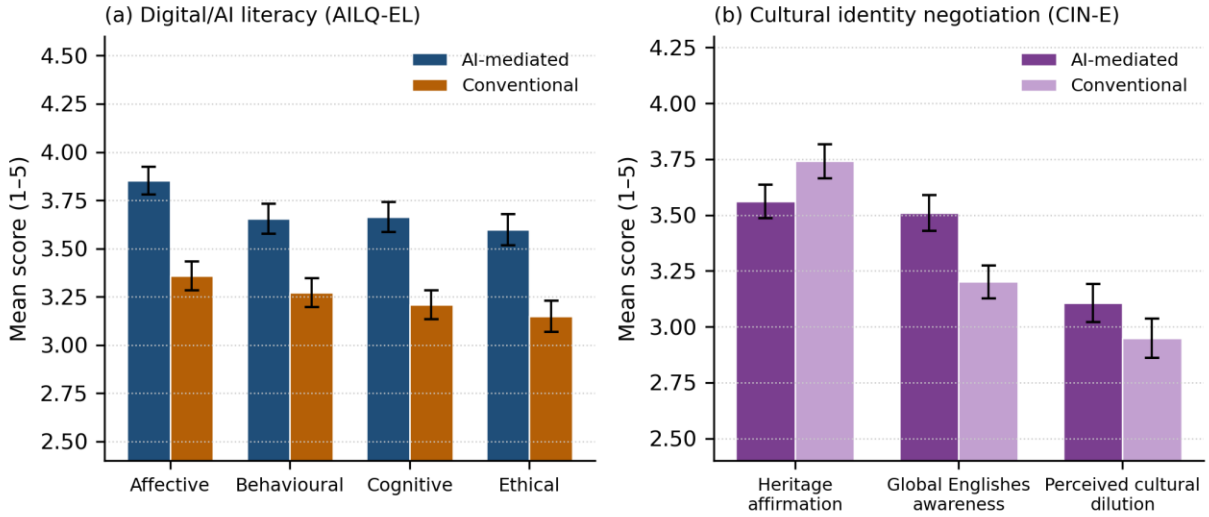

Figure 5. Effects of AI-mediated instruction on (a) the four dimensions of digital/AI literacy and (b) the three dimensions of cultural identity negotiation.

3.8 Associations among the constructs

The correlation matrix is presented in Table 8. The highest correlation amongst the three with post test attainment was with engagement ($r = 0.63$) and pre-test score ($r = 0.57$), followed by digital/AI literacy ($r = 0.56$). Engagement and literacy were also found to be fairly strongly correlated ($r = 0.59$), providing support for a serial mediation (as opposed to parallel mediation) specification (see Section 3.9). There was a modest individual level correlation between the objective IEQ and engagement ($r = 0.44$) and attainment ($r = 0.31$), and more strongly at the section level, where it is appropriately measured (Section 3.9). Students who felt their voice being flattened did not, on average, write worse essays, with hardly any correlation with cultural dilution ($r = -0.04$, $p = .326$). That sort of disconnect is worth noting, as it demonstrates that the cultural costs of AI mediation are not captured by a measure of attainment.

Table 8 *Bivariate correlations among study variables*

Variable	1	2	3	4	5	6	7	8	9
1. Digital/AI literacy	—								
2. Engagement	.59**	—							
3. Heritage affirmation	.07	.12**	—						
4. GE awareness	.27**	.24**	.08	—					
5. Cultural dilution	-.04	-.05	.01	-.00	—				
6. Spatial affordance	.15**	.40**	.04	.04	-.04	—			
7. IEQ index	.19**	.44**	.05	.04	-.04	.92**	—		
8. LIC pre-test	.28**	.28**	.14**	.17**	-.08*	-.05	-.05	—	
9. LIC post-test	.56**	.63**	.09*	.23**	-.04	.28**	.31**	.57**	—

To quantify the contribution of each layer of the model a four block hierarchical regression was used (Table 9). The model accounted for 33.8% of the variance in the post-test scores, with pre-test score and demographic covariates jointly accounting for the variance. Adding the two experimental factors raised this by 22.1 percentage points, $\Delta F(2, 589) = 147.38$, $p < .001$. The addition of these process variables (digital/AI literacy, engagement and perceived spatial affordance) accounted for an additional 7.5 points, $\Delta F(3, 586) = 39.86$, $p < .001$, and rendered the instruction coefficient statistically insignificant ($b = 0.617$, $p = 0.074$) in favor of the mediation model, the hallmark of mediation. The interaction term entered in Block 4 remained significant after all of this, $b = 1.996$, $\Delta R^2 = 0.011$, $\Delta F(1, 585) = 17.47$, $p < .001$. The final model explained 64.4% of the variance (adjusted $R^2 = 0.637$).

Table 9*Hierarchical regression predicting post-test literary interpretive competence (final model)*

Predictor	b	SE	β	t	p
Pre-test LIC score	0.581	0.033	0.486	17.45	< .001
Age	-0.144	0.080	-0.045	-1.79	0.074
Gender (male = 1)	0.143	0.236	—	0.60	0.547
AI-mediated instruction	0.617	0.345	0.066	1.79	0.074
Adaptive Learning Studio	0.903	0.607	0.096	1.49	0.137
Digital/AI literacy	1.563	0.341	0.154	4.58	< .001
Student engagement	2.197	0.343	0.234	6.40	< .001
Perceived spatial affordance	-0.047	0.261	-0.011	-0.18	0.858
AI × Studio interaction	1.996	0.478	—	4.18	< .001

The attainment model was re-estimated as a linear mixed model with a random section intercept as sections were randomised, not students. The unconditional intraclass correlation was 0.232, providing evidence that there was a non-trivial proportion of variance between sections and that the single level standard errors would be anti-conservative. Estimates from the conditional model are given in Table 10. The substantive results do not change: the interaction term is the largest condition effect ($b = 2.021$, 95% CI [1.013, 3.029], $p < .001$), and the instruction main effect is accounted for by the mediators ($b = 0.657$, $p = 0.076$). The model

accounted for 63.9% of the total variance as compared to the null model.

Table 10 *Linear mixed model predicting post-test literary interpretive competence, with random section intercepts*

Effect	b	SE	95% CI	z	p
Intercept	23.197	0.277	[22.655, 23.739]	83.88	< .001
Pre-test LIC (centred)	0.585	0.034	[0.519, 0.651]	17.36	< .001
AI-mediated instruction	0.657	0.370	[-0.069, 1.383]	1.77	0.076
Adaptive Learning Studio	0.849	0.368	[0.128, 1.570]	2.31	0.021
AI × Studio	2.021	0.514	[1.013, 3.029]	3.93	< .001
Digital/AI literacy (centred)	1.563	0.345	[0.886, 2.239]	4.53	< .001
Engagement (centred)	2.142	0.346	[1.465, 2.819]	6.20	< .001
Random effects					
Section intercept variance	0.054				
Unconditional ICC	0.232				

3.9 H6: Mediation and moderated mediation

H6 proposed that AI mediation will improve attainment by first improving digital/AI literacy which will lead to improved engagement. This is corroborated by the serial model (Figure 6, Table 11). AI-mediated instruction predicted digital/AI literacy (a_1 = 0.446, 95% CI [0.383, 0.513]) and, less strongly, engagement directly (a_2 = 0.124 [0.05, 0.2]). Literacy strongly predicted engagement (d_{21} = 0.576 [0.496, 0.655]). Both mediators predicted attainment (b_1 = 1.648 [0.883, 2.419]; b_2 = 3.44 [2.826, 4.076]). All three indirect routes were important but not all equal. The serial path, AI → literacy → engagement → attainment, was the largest at 0.883 points [0.667, 1.137], followed by the literacy-only path (0.735 [0.377, 1.119]) and the engagement-only path (0.427 [0.164, 0.71]). The indirect effect was 2.045 [1.617, 2.515] and the total effect was 3.207, which means that about 64% of the instructional effect was mediated by these two mediators. The direct effect did not fully account for the mediation effect, as this was still significant (c' = 1.079 [0.495, 1.644]), suggesting partial mediation and therefore a possibility of unmeasured pathways. The pedagogical order is of importance. If the engagement-only path had prevailed, the logical assumption would be that AI is successful because students like it. The dominance of the serial path, however, indicates that AI is effective because it develops a skill — being able to assess and challenge the generated text — that then sustains engagement, which then underpinned attainment. We emphasize that this is only a cross-sectional mediation analysis in which mediators were assessed once immediately after the intervention. The model assumes the ordering of time; it is not derived from the design.

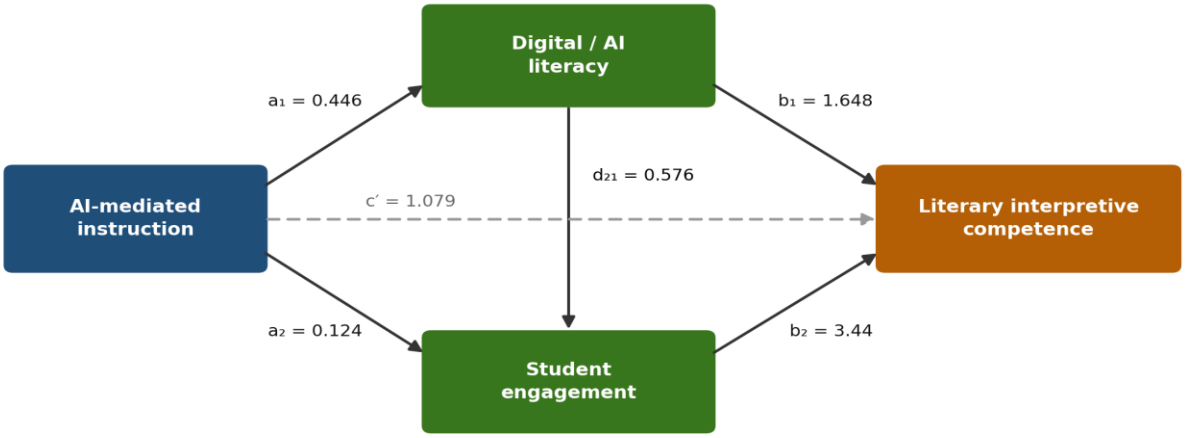

Figure 6. Serial mediation model of the effect of AI-mediated instruction on literary interpretive competence through digital/AI literacy and student engagement.

Table 11 *Serial mediation of the effect of AI-mediated instruction on literary interpretive competence*

Effect	Estimate	Boot SE	95% bootstrap CI	Inference
Total effect (c)	3.207	0.287	—	< .001
Direct effect (c')	1.079	0.297	[0.495, 1.644]	< .001
Indirect effects				
Via digital/AI literacy only	0.735	0.188	[0.377, 1.119]	Significant
Via engagement only	0.427	0.139	[0.164, 0.710]	Significant
Serial: literacy → engagement	0.883	0.120	[0.667, 1.137]	Significant
Total indirect	2.045	0.229	[1.617, 2.515]	Significant

The second part of H6 believed that this indirect path should also be contingent to the room. This was supported by a moderated mediation model that found the environment was a moderator of the engagement→attainment path (Table 12). The conditional indirect effect was 0.882 [0.538, 1.23] in conventional rooms and 1.403 [0.966, 1.878] in studios, giving an index of moderated mediation of 0.52 [0.097, 0.99]. The claim for moderated-mediation is supported with less confidence than the claim for simple interaction reported in Section 3.4, because the lower bound is near zero.

Table 12 *Moderated mediation: architectural environment as moderator of the engagement–attainment path*

Path or effect	Estimate	95% bootstrap CI	CI excludes 0
AI → engagement (a)	0.381	[0.307, 0.456]	Yes
Engagement → attainment in CLR (b)	2.317	[1.469, 3.108]	Yes
Difference in b between environments	1.367	[0.276, 2.472]	Yes
Conditional indirect effect, CLR	0.882	[0.538, 1.230]	Yes
Conditional indirect effect, ALS	1.403	[0.966, 1.878]	Yes
Index of moderated mediation	0.520	[0.097, 0.990]	Yes

Since IEQ is a property of rooms, rather than students, the environmental association was also examined on the section level, k = 24. Section mean engagement was correlated with the objective IEQ composite with r = 0.661, p < .001 and section mean attainment was correlated with the objective IEQ composite with r = 0.601, p = 0.002 (Figure 7). These correlations are far from certain because there are only 24 clusters, and the IEQ index is bimodal, arising from the fact that the index has two room types instead of a continuum, and should not be interpreted as a linear dose–response relationship. What they do find is that the impact of "the room" follows the actual measured characteristics of that room, rather than just its

name.

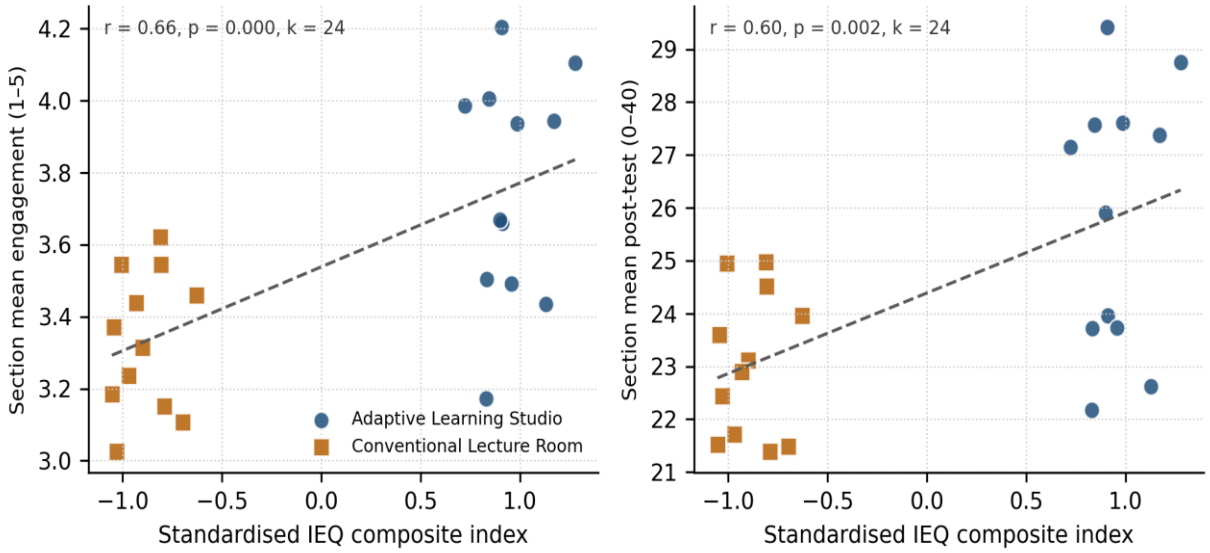

Figure 7. Section-level association between the objective indoor environmental quality composite and (a) mean engagement and (b) mean post-test attainment.

Note. Each point is one class section (k = 24). The dashed line is the ordinary least squares fit across all sections. The gap along the x-axis reflects the two engineered room types and precludes inference about intermediate IEQ values.

Figure 8 summarises all standardised contrasts on a common scale, allowing the instructional and environmental effects to be compared directly. Three patterns are visible at a glance: the instructional effect dominates for digital/AI literacy, the environmental effect dominates for agentic engagement, and the two are of comparable size for attainment and composite engagement. The cultural outcomes sit close to zero for the environmental contrast throughout.

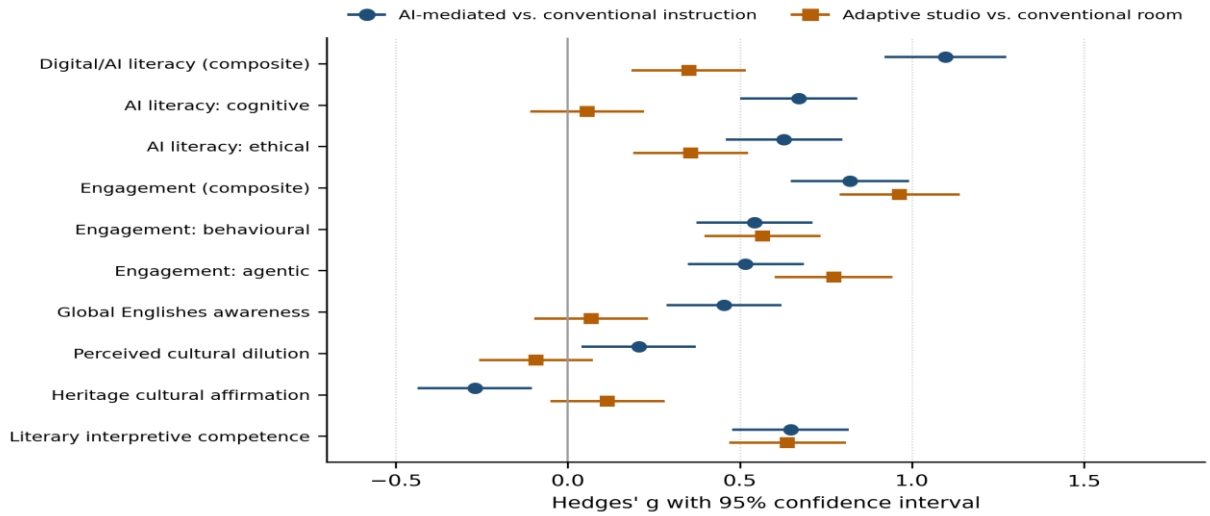

Figure 8. Standardised effect sizes (Hedges' g) with 95% confidence intervals for the instructional and environmental contrasts across all outcomes.

Note. Positive values indicate higher scores under AI-mediated instruction (circles) or in the Adaptive Learning Studio (squares). Intervals crossing zero indicate contrasts that were not statistically significant.

3.10 Qualitative findings and integration

Seven themes emerged from the thematic analysis of 42 student interviews and 12 instructor interviews, through a reflexive

approach (Table 13, Figure 9). They are shown in the order that they relate to the quantitative results, not in order of frequency. Cognitive offloading vs interpretive ownership (36/42). The prime tension that students reported was between the speed of an AI-generated reading and the sense of ownership of that reading. In all cases students in the AIM condition reported an early phase of accepting uncritically and a learned scepticism they said was related to the annotation procedure. Teachers explained the same trajectory from the other side. This theme resonates with the quantitative result that the association of protected cultural voice was not with the behavioural subscale, but with the ethical and cognitive literacy subscales. EUC (31/42) Register homogenisation and loss of authorial voice. Most AI-condition students said they had seen their writing become more standardized; a few explicitly spoke of actively resisting the machine in an effort to regain a personal voice. This is a triangulation with the relatively small increase in the perceived cultural dilution scale, and provides a phenomenological account of what the perceived scale is measuring. The literary exemplars are generated by the AI and do not align with the cultural context of the students (29/42). Students repeatedly reported that AI output was fluent when reading standard British and American English, but either lacked fluency or flattened Urdu literary reference, local idioms and expressions. Some reported that this mismatch was instructive when it became visible, as it was on average negative in the quantitative pattern — a mismatch that might account for the shift in awareness of Global Englishes at the same time as dropping in heritage affirmation was reported. Spatial reconfiguration that allows for peer critique (27/42). Studio-based participants reported that furniture movers were a regular part of the session and that they could look at the work of several other groups at the same time, but conventional-room participants reported that the furniture movers were an effortful task and were often stopped before the furniture would be moved. Teachers of the traditional rooms said that they squeezed or shortened the stage of display and critique due to the time constraint. This is the most direct qualitative description of the interaction that follows the quantitative interaction, and indicates that the "environment effect" can be part of the "implementation-fidelity effect", mediated by the room. Increase in confidence within participation (34/42) and equity constraints (25/42). "Confidence" gains were reported by a large number of people, and this was explained by the low stakes of practicing an argument with a machine first, before actually speaking it out loud. Opposed to this, a quarter of the interviewees stated that connectivity issues, shared-device restrictions and the costs associated with data usage were responsible for the failure, especially in the rural localities. Students who had only a smartphone gave a detailed description of the problem with annotate-and-compare. It is an important boundary condition of any recommendation that comes out of the study. Questions about proper use (23/42). In spite of the required attribution log, more than half of the interviewees still felt unsure about the boundaries of acceptable assistance. A few teachers expressed doubt in their own assessment decisions.

Table 13 *Qualitative themes, frequency and relationship to the quantitative findings*

Theme	n/42	%	Relationship to quantitative strand
Cognitive offloading vs. interpretive ownership	36/42	85.7%	Converges with ethical-literacy correlation (Section 3.7)
Register homogenisation and loss of authorial voice	31/42	73.8%	Converges with rise in perceived cultural dilution
Confidence gains in oral and written participation	34/42	81.0%	Converges with engagement gains; no quantitative counterpart for confidence
Cultural mismatch in AI-generated literary exemplars	29/42	69.0%	Diverges: mismatch reported as instructive as well as costly
Spatial reconfiguration enabling peer critique	27/42	64.3%	Explains the instruction × environment interaction

Connectivity, device and equity constraints	25/42	59.5%	Boundary condition; not captured by the outcome measures
Uncertainty about legitimate use and attribution	23/42	54.8%	Expands on the ethical subscale; no quantitative counterpart

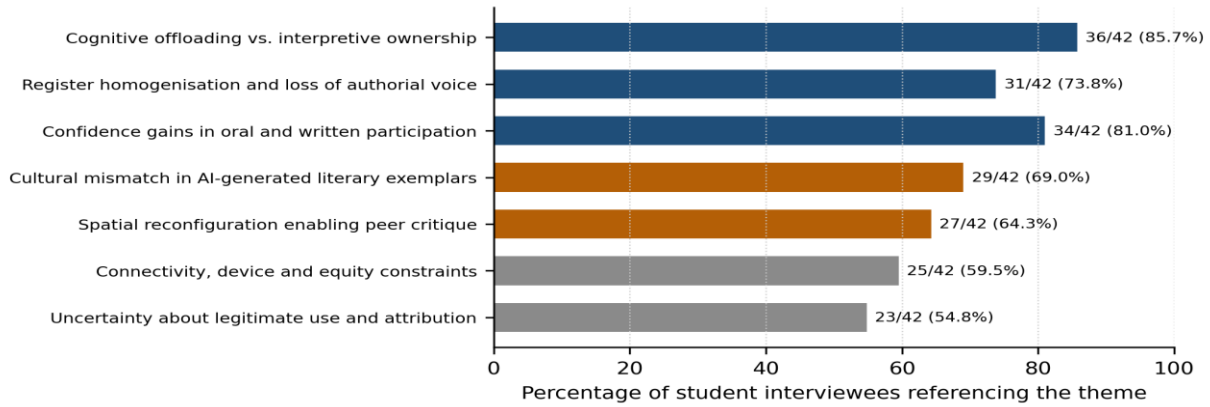

Figure 9. Frequency of qualitative themes across 42 student interviews.

3.11 Summary of hypothesis tests

Table 14 consolidates the evidential status of each hypothesis and, importantly, distinguishes claims supported by the experimental manipulation from claims resting on measured covariation. Only the first category licenses causal language.

Table 14 *Evidential status of each hypothesis*

Hypothesis	Verdict	Basis of inference	Key statistic
H1: AI raises attainment	Supported	Experimental (randomised)	$\eta^2p = 0.216$, $g = 0.647$
H2: AI raises engagement (all dimensions)	Supported	Experimental (randomised)	$\eta^2p = 0.182$, all subscales $p < .001$
H3: AI raises digital/AI literacy, largest on cognitive and ethical	Partially supported	Experimental (randomised)	Direction confirmed; predicted ordering not obtained
H4: AI raises GE awareness, lowers heritage affirmation, raises dilution	Supported, small effects	Experimental for group means; associational for mechanism	$g = 0.454$, -0.27, 0.206
H5: Studio raises outcomes and amplifies the AI effect	Supported	Experimental (randomised)	Interaction $\eta^2p = 0.059$, $p < .001$
H6: Serial mediation, conditional on environment	Supported with caution	Associational (concurrent mediators)	Total indirect = 2.045; index of moderated mediation = 0.52
Exploratory: ethical literacy buffers dilution	Consistent, untested	Associational, unplanned	$r = -0.296$, $p < .001$; tercile contrast n.s.

4. Discussion

4.1 Principal findings

The aim of this study was to determine if the effects of AI-supported English language and literature instruction are contingent on the design of the learning space in which it takes place and how it affects learners' own cultural voice. There are four noteworthy findings.

AI mediation brought success; and the most successful result was in a near-to-the-point outcome. Results demonstrated that the largest instructional effect in the study was for digital and AI literacy ($\eta^2p = 0.241$, $g = 1.097$), followed by engagement ($\eta^2p = 0.182$) and interpretive competence ($\eta^2p = 0.216$). The magnitude and ordering is similar to that found in the meta-analytic research record: Liu et al. (2025) and Wang and Fan (2025) both found positive average effects of ChatGPT on achievement and significant heterogeneity, while Wu and Li (2024) found similar effects for chatbot-supported skill development. The attainment effect observed here is not at the rosy end of that range but in the middle; that's what an active-control design should yield. As this comparison focused solely on the collaborative close-reading cycle and omitted the AI component, it also filters out the synergy between AI and the more interactive pedagogy, which Lo et al. (2024) refer to as a "confound" that is common in this body of research.

Second, but most importantly, the effect of the instruction was dependent on the room. The interaction effects of instruction × environment were significant for attainment ($\eta^2p = 0.059$), and for three of the four dimensions of engagement, and differed by a factor of 2.8 ($g = 0.954$ in the Adaptive Learning Studio and $g = 0.4$ in the Conventional Lecture Room) with non-overlapping confidence intervals. The integration of AI in a fixed-row lecture room generated about the same effect as the improvement of the room, with no effect on pedagogy, while the integration of the two together generated more than the sum of its participants. To our best knowledge, this is the first direct cross-examination to be carried out to prove that the effect of the integration of AI in language learning is moderated by the architectural setting. A concrete and testable explanation for part of the heterogeneity that is a problem for existing syntheses: the former study is done in reconfigurable spaces, and the latter in lecture halls; each of these usually does not report which it was.

Thirdly, motivation seems to be a feature of capability and not of enthusiasm. The largest of the three indirect paths was the serial pathway (AI → digital literacy → engagement → attainment; 0.883 points, 95% CI [0.667, 1.137]) which surpassed the path through engagement only (0.427). If the opposite was true, the natural reading would be that the use of AI increases attainment because students enjoy it, which is something that can be easily lost due to novelty decay. The observed ordering, however, points to the evaluative competence that students develop by contesting machine output which also maintains engagement of the durable asset. This is consistent with the work of Cardon et al. (2023), who suggest that the competence being shown in the use of AI in writing is judgement about when to reject a suggestion; and with the work of Delcker et al. (2024), who find that low-level AI competence predicted future use, thus indicating a self-reinforcing capability loop.

Fourth, the cultural results were truly two-sided, and are the ones we would most desire to see repeated. The awareness of global English increased ($g = 0.454$), the affirmation of heritage culture decreased ($g = -0.27$), and perceived cultural dilution increased ($g = 0.206$) under the mediation of AI. Their impact is small and their dilution interval is at its lower limit equal to zero; it's a shift that you can detect, not an identity crisis.

4.2 The cultural tension, and what the qualitative data add

The critical literature has suggested on text-based evidence that the native-speaker English eschewed by LLM are standardised Englishes, which de-emphasise minority varieties (Dai et al., 2025; Jenks, 2025; Jeon et al., 2025). At the model output level, Lee et al (2025) showed that an unmodified ChatGPT was unable to produce output that was aligned with Korean English, while the inclusion of retrieval-augmented, corpus-grounded customisation helped to improve the situation. What was missing is evidence on the learner, not models. The present data provide the first rough quantitative estimate, and it is in agreement with the criticism in direction, though it is significantly smaller in

numerical strength than the criticism in that direction would imply.

Both can be true, as explained by these interviews. The students reported that they were aware of the tendency for their prose to become more and more similar to each other, and that they consciously and deliberately worked to achieve "against" the machine to regain their personal style (31 of 42 respondents). However, once the mismatch between AI output and South Asian English writing became obvious, 29 of 42 responded saying that it was "instructive". On this score, AI is used as a sort of "contrast medium" – it makes the default-English/community-English readable, precisely in not being able to accept local literary reference. That's why the otherwise baffling mixture of increased awareness and decreased affirmation makes perfect sense, and why we do not read the affirmation decline as so devastating simply as a drop in affirmation. It can also be a positive destabilisation of an unchallenged stance. The present design does not identify these interpretations, but we note the ambiguity.

The only significant correlation is between ethical AI literacy and protected voice ($r = -0.296$, $p < .001$), but not between the two and usage-oriented literacy ($r = .09$, $p = .121$). This is a nonplanned correlational finding in one arm of the study only, and should not be interpreted as a protective effect. It is, however, aligned with the multidimensional view of AI literacy that has been formulated by Ng et al. (2021, 2024) and with instrument development efforts, which have distinguished critical appraisal from practical activity (see Hornberger et al., 2023; Laupichler et al., 2023). It also suggests that an AI-literacy definition limited to training as 'a tool' – the framing that is most commonly used in institutional guidance – is likely to be the most likely to fail to defend cultural interests.

4.3 The importance of the room, and the importance of which part of the room

The environmental effect was not uniform over outcomes and its shape was informative. The largest was for agentic engagement ($\eta^2p = 0.14$, $g = 0.772$), and the smallest was for cognitive engagement ($\eta^2p = 0.071$), with no detectable effect on any cultural outcome. A room that allows students to turn, move and claim a display surface has an impact on the dimension of engagement defined by initiative and ownership, but little on the privacy of textual processing, and nothing measurable on cultural voice. A general positive or demand-characteristic explanation would assume that there would be a constant increase in elevation across all outcomes, but this was not the case.

This design is such that two aspects of the manipulation are confounded and cannot be separated here. Objective IEQ was significantly higher for some room types than others with respect to each of the measured parameters, and the correlation between the IEQ composite and mean engagement level was large across the sections ($r = 0.661$, $p < .001$, $k = 24$). This is in line with Brink et al. (2021) who, through their systematic review, found that IEQ had a positive effect on short-term academic performance in higher education and was in line with the field experiment of Brink et al. (2023) that found better perceptions and problem-solving in higher education classrooms due to illuminance improvement and reverberation reduction to class-A standards. However, the studios were also, as the qualitative strand reports, configured differently and were easier to reconfigure, which the qualitative strand marks as the more immediate cause: Instructors in more conventional studios reported that they usually had to compress or abandon the display and critique phase within the space, consistent with Parsons (2024) discussion of the agency supporting properties of non-frontal arrangements and Bingen et al. (2023) discussion of re-imagining classrooms for active learning. For EFL in particular, Ye (2024) shows similar findings of the relationship between perceived classroom climate and engagement, and Bizimana (2025) shows similar associations between perceived classroom environment with engagement. Part of what we've termed the environmental effect is, in turn, plausibly an implementation-fidelity effect transmitted through the room, i.e. whether the intended pedagogy occurs or not. The most obvious next study is to disentangle ambient quality from spatial affordance, as this requires a design that transgresses them.

4.4 Theoretical implications

The findings suggest a combined intervention of integrating the use of AI and designing learning spaces rather than viewing them as separate interventions. From a sociocultural perspective this is the expected outcome: If AI is a mediating

artefact that reorganises an activity system, then its impact is contingent on the other components of the activity system, such as its material environment. The contribution here is to assign a number to that conditionality. The view of self-determination suggests that the two manipulations offer different psychological nutrients: competence feedback from the machine, autonomy and relatedness from the room, which explains the super-additive effect of the two manipulations when combined. The current disengagement contributes to sharpening engagement, which has long been under-theorised in this field, as pointed out by systematic evidence maps of technology and engagement (Bond et al., 2020).

The results suggest that the cultural outcomes should be part of the effectiveness model instead of an ethics appendix for globalizing English. The only significant link between cultural dilution and attainment was non-existent (r = -0.04), meaning that cultural dilution is invisible to any evaluation system that is based on grades. This would prove to an institution, who only looks at attainment, to be a complete success.

4.5 Practical implications

The implications are as follows:

Plan AI and space at the same time. The data imply that whilst these data indicate that introducing AI in unaltered lecture rooms will realise a small proportion of the benefit, where resources are required to make a decision. In addition to the technology, there are a number of modest modifications to rooms which are less expensive than most institutional technology procurement; these modifications can make a difference, as much as the technology, especially in this instance.

Educate ethical and evaluative AI skills rather than tool skills. The association with protected cultural voice occurred in the ethical subscale, rather than the usage subscale, and the cognitive and ethical dimensions led the serial mediation path.

Include cultural failures in the curriculum. It was found that students were still learning from AI's mishandling of South Asian literary reference when it was brought to their attention.When explicitly brought to students' attention, it was found that students continued to learn from AI's mishandling of South Asian literary reference. If corpus-grounded customisation is not possible, then the failure itself can be taught, as demonstrated by Lee et al. (2025) in the context of improving model behaviour.

Audit for equity before scaling. A quarter of interviewees mentioned connectivity and device/ data-cost as constraints, with the majority of these respondents being from rural regions and smartphone-only students. Ashraf et al. (2025) find the same level of access disparities in the Pakistani higher education sector. An overall average attainment gain could mask an actual widening gap.

4.6 Limitations

These were several restrictions that limited these conclusions, and it is not our intention to mince words about them.

The design used is quasi-experimental and it is at individual level. Randomisation was by intact section; this is suitable for an environmental manipulation, but pre-existing differences between sections could exist; the unconditional intraclass correlation of 0.232 further indicates that there was a pre-existing difference due to section membership. Having 24 clusters, the power at the cluster level is small, and the section-level correlations in Section 3.9 are not precise.

The mediation analyses are correlational. The mediators were measured at the same time as the outcome was measured at Week 16, so the order shown in Figure 6 is adopted rather than designed. The covariance matrix can be used with both the reverse and the reciprocal orderings, of which the first is attainment raising engagement and the second is engagement raising reported literacy. The lower bound (0.097) of the moderated-mediation index (0.52 [0.097, 0.99]) is near zero and is therefore suggestive.

Three of the four constructs are based on self-report, which introduces potential social desirability and common-method variance, but Harman's test and scalar invariance testing did not suggest that this was a serious problem. The CIN-E instrument is new, its factorial validity has been demonstrated here but not its construct validity with behaviour indicators of cultural voice (such as the actual use of local literary reference in writing, for example).

As mentioned above, the environmental manipulation is confounded with ambient quality; the perceived-affordance measure is also

highly correlated with the room dummy (VIF = 5.65) and thus cannot be interpreted separately from the room dummy. There isn't much here to suggest that the attainment gain will last, that the cultural change will be strengthened or weakened, or that engagement gains will outlast the novelty. Short intervention windows is one of the most common weaknesses reported in educational reviews of ChatGPT (Montenegro-Rueda et al., 2023), which is also the case here.

Lastly, there is limited generalisability. Each of the four sites is a public sector university in one province of Pakistan, whose students learn from a common curriculum, and whose language of instruction is English, with a special postcolonial connotation for them. It is not advisable to extrapolate the cultural findings to contexts in which the charge varies, and comparative research involving the contexts of other sociolinguistic settings, similar to what was done in Klimova et al. (2024), Xiao and Zhi (2023) and Yu et al. (2025) would be needed before any such claim.

We also don't state something that we can't prove with the data. There is no evidence here that AI teaching enhances overall critical thinking or writing skills beyond the rubric used for assessment, or that AI teaching has a positive long-term impact. This results in a 40-point analytic essay grade for an unseen poem – this is the only attainable claim we make.

4.7 Directions for future research

Four studies would make a difference to this research stream. A 2x2 Experimental design, crossing ambient IEQ by spatial reconfigurability, would split up the two parts of the environmental effect. Instead of assuming a temporal ordering, a longitudinal design with mediators measured at three or more time points would test it; and it would determine whether engagement gains decrease over time. The correlational buffering result would become a causal test if a randomized comparison of tool-focused versus ethics-focused AI-literacy instruction was made. Finally, a design comparing the default LLM output to retrieval-augmented, locally corpus-grounded models, in line with the customisation approach of Lee et al. (2025), would allow to test directly the cultural cost observed here, as a property of AI mediation in general, or of the specific models being used in classrooms. Some of these are provided by the instrument on work done with regard to learner perceptions and adoption (Derakhshan & Ghiasvand, 2024; Jeon & Lee, 2023; Sandu et al., 2024; Song & Song, 2023), and on intention-to-use measurement (Chai et al., 2024).

5. Conclusion

In this sample of 597 undergraduates, teaching through artificial intelligence in English language and literature increased digital and AI literacy and engagement and literary interpretive competence in the moderate-to-large range. The more consequential result is that these benefits were dependent on the room: the instructional effect was 2.8 times greater in a reconfigurable, well-lit, acoustically treated studio, compared to a traditional lecture room, and the interaction remained after controlling for pre-test performance, clustering, and measured mediators. Architecture is not a background factor in educational technology research, but rather it is a moderator that has an effect size similar to the technology itself.

Any simple praise is tempered by the cultural results. AI mediation increased students' awareness of English as a plural, contested resource, and resulted in minor reductions in heritage cultural affirmation and minor increases in a sense that English written voice was being flattened. These changes were not attributable to attainment, therefore an institution looking at grades alone would not see these changes when considering this intervention. The correlation of ethical AI literacy and protected voice points towards a solution, but not a solution that has been proven in the wild yet, it is a correlation within one arm of one study, but one that should be experimented upon in order to be implemented on a wider scale.

The practical message is the narrow, but hopefully useful, one that institutions introducing generative AI into humanities teaching should consider the technology and the classroom teaching space as one intervention, that they should teach critical and ethical AI literacy instead of tool literacy, and that they should try to measure the effects of the intervention on the cultural voice of students, as they will never be told by attainment data.

References

Alhusaiyan, E. (2025). A systematic review of current trends in artificial intelligence in foreign language learning. *Saudi Journal of Language Studies, 5*(1), 1–16. https://doi.org/10.1108/SJLS-07-2024-0039

Ashraf, M. A., Alam, J., & Kalim, U. (2025). Effects of ChatGPT on students' academic performance in Pakistan higher education classrooms. *Scientific Reports, 15*(1), Article 16434. https://doi.org/10.1038/s41598-025-92625-1

Bingen, H. M., Aamlid, H. I., Hovland, B. M., Nes, A. A. G., Larsen, M. H., Skedsmo, K., Petersen, E. K., & Steindal, S. A. (2023). Use of active learning classrooms in health professional education: A scoping review. *International Journal of Nursing Studies Advances, 5*, Article 100167. https://doi.org/10.1016/j.ijnsa.2023.100167

Bizimana, E. (2025). Exploring the contribution of perceived supportive classroom learning environment to students' engagement in learning. *European Journal of Psychology and Educational Research, 8*(2), 97–112. https://doi.org/10.12973/ejper.8.2.97

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education, 17*, Article 2. https://doi.org/10.1186/s41239-019-0176-8

Brink, H. W., Krijnen, W. P., Loomans, M. G. L. C., Mobach, M. P., & Kort, H. S. M. (2023). Positive effects of indoor environmental conditions on students and their performance in higher education classrooms: A between-groups experiment. *Science of the Total Environment, 869*, Article 161813. https://doi.org/10.1016/j.scitotenv.2023.161813

Brink, H. W., Loomans, M. G. L. C., Mobach, M. P., & Kort, H. S. M. (2021). Classrooms' indoor environmental conditions affecting the academic achievement of students and teachers in higher education: A systematic literature review. *Indoor Air, 31*(2), 405–425. https://doi.org/10.1111/ina.12745

Brink, H. W., Loomans, M. G. L. C., Mobach, M. P., & Kort, H. S. M. (2022). A systematic approach to quantify the influence of indoor environmental parameters on students' perceptions, responses, and short-term academic performance. *Indoor Air, 32*(10), Article e13116. https://doi.org/10.1111/ina.13116

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly, 86*(3), 257–295. https://doi.org/10.1177/23294906231176517

Celik, I. (2023). Exploring the determinants of artificial intelligence (AI) literacy: Digital divide, computational thinking, cognitive absorption. *Telematics and Informatics, 83*, Article 102026. https://doi.org/10.1016/j.tele.2023.102026

Chai, C. S., Yu, D., King, R. B., & Zhou, Y. (2024). Development and validation of the artificial intelligence learning intention scale (AILIS) for university students. *SAGE Open, 14*(2). https://doi.org/10.1177/21582440241242188

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology, 55*(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Dai, D. W., Suzuki, S., & Chen, G. (2025). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review, 16*(2), 763–774. https://doi.org/10.1515/applirev-2024-0184

Deep, P. D., Martirosyan, N., Ghosh, N., & Rahaman, M. S. (2025). ChatGPT in ESL higher education: Enhancing writing, engagement, and learning outcomes. *Information, 16*(4), Article 316. https://doi.org/10.3390/info16040316

Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students' AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *International Journal of Educational Technology in Higher Education, 21*, Article 18. https://doi.org/10.1186/s41239-024-00452-7

Derakhshan, A. (2025). EFL students' perceptions about the role of generative artificial intelligence (GAI)-mediated instruction in their emotional engagement and goal orientation: A motivational climate theory (MCT) perspective in focus. *Learning and Motivation, 90*, Article 102114. https://doi.org/10.1016/j.lmot.2025.102114

Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions. *International Journal of Applied Linguistics, 34*(4), 1246–1264. https://doi.org/10.1111/ijal.12561

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, Article 100165. https://doi.org/10.1016/j.caeai.2023.100165

Jenks, C. J. (2025). Communicating the cultural other: Trust and bias in generative AI and large language models. *Applied Linguistics Review, 16*(2), 787–795. https://doi.org/10.1515/applirev-2024-0196

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Jeon, J., Li, W., Tai, K. W. H., & Lee, S. (2025). Generative AI and its dilemmas: Exploring AI from a translanguaging perspective. *Applied Linguistics, 46*(4), 709–717. https://doi.org/10.1093/applin/amaf049

Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology, 15*, Article 1269319. https://doi.org/10.3389/fpsyg.2024.1269319

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy" (SNAIL): An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, Article 100338. https://doi.org/10.1016/j.chbr.2023.100338

Lee, S., Jeon, J., McKinley, J., & Rose, H. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics, 45*, 85–108. https://doi.org/10.1017/S0267190525100184

Liu, Z., Zuo, H., & Lu, Y. (2025). The impact of ChatGPT on students' academic achievement: A meta-analysis. *Journal of Computer Assisted Learning, 41*(4), Article e70096.

Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education, 219*, Article 105100. https://doi.org/10.1016/j.compedu.2024.105100

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers, 12*(8), Article 153. https://doi.org/10.3390/computers12080153

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence, 3*, Article 100054. https://doi.org/10.1016/j.caeai.2022.100054

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Parsons, C. S. (2024). Erasing the line between students and instructors: The influence of classrooms and online learning spaces on student engagement. *Journal of the Scholarship of Teaching and Learning, 24*(4).

Sandu, R., Gide, E., & Elkhodr, M. (2024). The role and impact of ChatGPT in educational practices: Insights from an Australian higher education case study. *Discover Education, 3*(1), 1–16. https://doi.org/10.1007/s44217-024-00126-6

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*(1), 1–21.

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2509759

Wu, X., & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *The Asia-Pacific Education Researcher*. Advance online publication. https://doi.org/10.1007/s40299-024-00853-2

Xia, Q., Zhang, P., Huang, W., & Chiu, T. K. F. (2025). The impact of generative AI on university students' learning outcomes via Bloom's taxonomy: A meta-analysis and pattern mining approach. *Asia Pacific Journal of Education*, 1–31.

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages, 8*(3), Article 212. https://doi.org/10.3390/languages8030212

Ye, X. (2024). A review of classroom environment on student engagement in English as a foreign language learning. *Frontiers in Education, 9*, Article 1415829. https://doi.org/10.3389/feduc.2024.1415829

Youssef, E., Medhat, M., Abdellatif, S., & Al Malek, M. (2024). Examining the effect of ChatGPT usage on students' academic learning and achievement: A survey-based study in Ajman, UAE. *Computers and Education: Artificial Intelligence, 7*, Article 100316. https://doi.org/10.1016/j.caeai.2024.100316

Yu, H., Guo, Y., Yang, H., Zhang, W., & Dong, Y. (2025). Can ChatGPT revolutionize language learning? Unveiling the power of AI in multilingual education through user insights and pedagogical impact. *European Journal of Education, 60*(1), Article e12749.