AI-Enhanced Outcome-Based College English Teaching Evidence of Initial Outcomes

Xiaoqing Zeng[1,2], Prabath Ekanayake[2]
[1]Shenyang City University, Shenyang, Liaoning, China
[2]University of Peradeniya, Peradeniya, Sri Lanka

Abstract:This paper reports the early outcomes of embedding artificial intelligence (AI) into an outcome-based College English course at Shenyang City University, where AI-supported practice and feedback were attached to tasks built around explicit learning outcomes. The analysis draws on two sources: a questionnaire from the 2024-entry cohort and the institution's course records for two consecutive stages, B-2 and B-3. Of 350 questionnaires, 329 were usable (94.00%); matching teaching-class names yielded 51 shared classes. The survey is encouraging: over 95% of students understood the course objectives, task requirements, and assessment criteria, OBE-item agreement ran 92–94%, and AI-support agreement 87–88%, with 95.74% wanting the approach to continue. The class records agree: across the 51 classes the weighted final-examination mean rose from 70.60 to 72.23 and the pass rate from 76.78% to 82.71%, with smaller gains in overall marks. The evidence does not establish causation, but it suggests that tying AI tools to explicit outcomes is a workable way to support College English students.

Keywords: Artificial Intelligence; Outcome-Based Education; College English; Teaching Reform; Learning Outcomes

1. Introduction

College English teaching in application-oriented undergraduate institutions is expected to address foundational language knowledge, examination preparation, and communication in practical contexts. These requirements call for explicit decisions about what students should be able to accomplish by the end of a course.Outcome-based education (OBE) organises teaching around intended student performance rather than covered content [1], and Biggs's constructive alignment argues that outcomes, activities, and assessment should reinforce one another so learners know the target and how it is judged [2]. In a language course this turns a vague "improve proficiency" into observable tasks—listening for gist, speaking to a purpose, reading critically, writing with structure.

Generative AI changes what individualised support can look like—explaining grammar, answering questions, giving writing feedback, and enabling out-of-class practice. The British Council's 2024 survey of teachers in 118 countries found AI used to support speaking, writing, and reading and to help students set goals, while noting that listening gets little attention and many teachers feel unprepared [3]; Kohnke, Moorhouse, and Zou [4] add that students need new digital literacies to use such tools well. The design question is not whether AI "helps" in the abstract, but how its functions can be attached to specific objectives and criteria so they extend, not bypass, intended learning.

This study examines that attachment in one setting. It pairs a student questionnaire with course records from two consecutive College English stages at Shenyang City University and asks three questions: how students actually use AI for English; how they judge the OBE design, the AI support, their own engagement, and perceived gains; and what shifts appear in cohort-level performance between B-2 and B-3. The aim is not a breakthrough but a stocktaking of early outcomes and the features worth keeping.

2. Theoretical Framework

For OBE in a language course, the starting point is the performance a learner should achieve, not the textbook chapter. Named performances then govern which activities and assessments make sense. Constructive alignment supplies the logic: when outcomes, teaching, and assessment point the same way, students better understand the target and how they will be judged [2]. The relevant outcomes here are concrete—comprehending information, organising ideas,

 Copyright @ STEMM Institute Press

interacting in real time, and carrying out language tasks in context.

Our frame sorts the course into four overlapping areas: clarity of goals and criteria, AI's contribution to the learning process, student engagement, and perceived gains. Clear requirements are the floor; AI then provides process support—explanations, practice, feedback. Engagement appears as participation, revision, and reflection, while perceived gains and actual marks are two lenses on the same question. Feedback research stresses that what matters is what the student does with the comment, not the comment itself [5]; we therefore attended to whether students acted on AI suggestions, revised substantively, and judged their own work.

In practice the arrangement shares responsibilities without removing anyone. Teachers set outcomes and criteria; AI handles explanation, practice, and first-pass suggestions; students do the talking, writing, and reflecting. UNESCO's guidance insists the technology serve a human-centred classroom rather than replace it [6]. We treat teacher guidance, intelligent support, and student agency as three mutually supporting strands. The frame is descriptive and makes no causal claim.

3. Research Design

3.1 Participants and Data Sources

The study combines a cross-sectional questionnaire with a comparison of existing course records, each answering a different part of the question. The survey reached across the university's 2024-entry cohort via Wenjuanxing; of 350 submissions, 21 were invalid, leaving 329 usable (94.00%). All respondents had taken both B-2 and B-3, allowing one survey to capture retrospective stage comparisons. Academic data came from the university's 2024-entry College English statistics: semester 2 of 2024–2025 (B-2) and semester 1 of 2025–2026 (B-3). Matching teaching-class names gave 51 classes taught in both stages, with 2,149 examination takers in B-2 and 2,001 in B-3.

3.2 Instruments

The questionnaire followed the four framework areas. Background items asked which AI tools students used (multiple response) and how often. The core statements were grouped as OBE implementation (B1–B5), AI-supported learning (C1–C7), engagement (D1–D5), perceived gains (E1–E6), and overall evaluation (F1–F4), rated on a five-point scale (1 = strongly disagree to 5 = strongly agree); reported “agree” totals combine the top two points. Some items offered a “not experienced / not applicable” option, so a few denominators differ. To map use onto activities, students rated how far each application helped them; the platform returned composite rankings showing relative emphasis, not time spent. Three open questions closed the survey and are used only to illustrate points already visible in the closed items.

3.3 Course Performance Comparison

Academic indicators were the final-examination mean and pass rate and the overall course mean and pass rate, weighted by examination participants per class; pass rates were recalculated from participant and failure counts. Because class sizes changed, a supplementary comparison held B-2 participant counts as fixed weights. Aggregate means are approximate (source means were pre-rounded). The comparison tracks cohort-level movement, not matched individual trajectories.

4. Results

The six subsections move from how students used AI, through their perceptions of the design and own learning, to what the institution's records show—from self-reported practice to independently recorded performance. The figures are in the tables; the commentary draws out what each block means.

4.1 AI Use Patterns and Learning Applications

The AI habit was already established: 157 students used it several times a week and 83 almost daily—240 (72.95%) at least a few times weekly; 83 (25.23%) used it occasionally and six (1.82%) never. General-purpose generative AI dominated tool use (290, 88.15%), followed by platform-embedded AI (205, 62.31%) and translation/writing aids (171, 51.98%); listening/reading/vocabulary apps (145, 44.07%), speaking tools (143, 43.47%), and content-creation tools (130, 39.51%) trailed but remained common (Table 1). In short, a versatile general tool did most of the work, with specialist aids in supporting roles.

By purpose, vocabulary and grammar ranked first (6.55), then reading (5.50), listening (4.25),

Copyright @ STEMM Institute Press

writing revision (3.88), spoken interaction (3.63), group projects (2.91), and reflection (1.69). Use thus concentrated on the language basics—words, rules, reading—while interactive and reflective uses lagged. The rankings show relative emphasis, not time spent.

Table 1. AI Tools Used for English Learning

Tool category	n	Percentage
General-purpose generative AI	290	88.15%
AI embedded in course platforms	205	62.31%
Translation and writing assistance	171	51.98%
Listening, reading, and vocabulary tools	145	44.07%
Speaking and pronunciation tools	143	43.47%
Content creation tools	130	39.51%

4.2 Clarity of Outcomes and Assessment Requirements

Students clearly registered the OBE design: 315 (95.74%) confirmed that teachers had explained the objectives, task requirements, and assessment criteria. Agreement across the five OBE items ran 92.10%–94.22% (means 4.48–4.53; Table 2), strongest on the link between activities and objectives and on criteria for improvement (both 94.22%). Most students could see how the course's pieces fitted together, and that visibility gave them a stable reference point. In effect, the objectives and criteria functioned as a frame students actively used, not background paperwork.

Table 2. Student Perceptions of OBE Implementation

Item	Agree n	Agree %	Mean
B1 Clear intended learning tasks	303	92.10%	4.49
B2 Activities linked to objectives	310	94.22%	4.53
B3 Authentic tasks supported application	303	92.10%	4.48
B4 Criteria guided improvement	310	94.22%	4.53
B5 Learning adjusted following feedback	304	92.40%	4.50

4.3 AI Process Support and Feedback-Based Revision

AI's process contribution was steady and uniformly positive. Agreement on the seven support items ran 86.63%–88.45% of the full sample and 92.53%–93.57% among those who had experienced them (means 4.43–4.50; Table 3); 17–22 students marked "not experienced" per item, reflecting uneven exposure rather than rejection. Whether the help was guidance on using AI, comprehension, spotting expression problems, or linking suggestions to criteria, evaluations stayed positive.

Table 3. Student Experiences of AI Support

Item	Valid n	Full sample %	Applicable %	Mean
C1 Guidance on AI use	308	86.63%	92.53%	4.49
C2 Understanding language and texts	307	86.93%	93.16%	4.47
C3 Identifying expression problems	309	87.84%	93.53%	4.45
C4 Substantive revision	309	87.54%	93.20%	4.43
C5 Checking accuracy	310	87.84%	93.23%	4.50
C6 Retaining personal viewpoints	312	87.84%	92.63%	4.43
C7 Alignment with objectives and criteria	311	88.45%	93.57%	4.46

The revision behaviour stands out: 288 (87.54%) made substantive changes after AI feedback, and 289 (87.84%) checked AI's accuracy and kept their own viewpoint. The reported use was thus judgement and rewriting, not passive consumption—AI entered at revision, where students weighed its suggestions.

4.4 Engagement and Perceived Learning Outcomes

Engagement was the norm: agreement ran 88.75%–92.40%, with proactive task completion and reflection tying highest (92.40% each), revision at 91.19%, and planning later study at 91.79% (Table 4). The pattern points to students completing, iterating on, and managing their own learning. On perceived gains, reading led (93.92%), then listening (92.10%), writing (90.58%), and speaking (88.75%); confidence on authentic tasks (92.10%) and critical AI use (91.79%) followed (means 4.28–4.35). The gains describe broad-based, incremental improvement across every skill and disposition—no single dramatic leap.

Table 4. Engagement and Perceived Learning Gains

Item	Agree n	Agree %	Mean
D1 Proactive task completion	304	92.40%	4.25
D2 Participation in	292	88.75%	4.24

 Copyright @ STEMM Institute Press

classroom interaction			
D3 Revision and improvement of work	300	91.19%	4.30
D4 Planning subsequent learning	302	91.79%	4.29
D5 Reflection and self-evaluation	304	92.40%	4.28
E1 Perceived listening improvement	303	92.10%	4.29
E2 Perceived speaking improvement	292	88.75%	4.28
E3 Perceived reading improvement	309	93.92%	4.35
E4 Perceived writing improvement	298	90.58%	4.33
E5 Confidence in authentic tasks	303	92.10%	4.32
E6 Critical use of AI	302	91.79%	4.34

4.5 Endorsement and Willingness to Continue

Support for the approach was near-unanimous yet not uncritical: 315 (95.74%) endorsed the AI–OBE blend and 308 (93.62%) wanted it to continue, while 89.36% worried about dependence and 88.75% had met inaccurate or unreliable feedback (Table 5). The concerns mark where teacher guidance and checking still matter; the net impression is an approach students want kept, but handled with care.

Table 5. Overall Evaluations and AI-Related Experiences

Item	Agree n	Agree %
F1 Endorsement of the teaching approach	315	95.74%
F2 Willingness to continue	308	93.62%
F3 Concern about reduced independent ability	294	89.36%
F4 Experience of inaccurate or unreliable feedback	292	88.75%

4.6 Improvement in Cohort-Level Academic Performance

The class records independently corroborate the survey. Every indicator was higher in B-3 than B-2 (Table 6): the weighted examination mean rose 1.63 points (70.60→72.23) and the pass rate 5.93 pp (76.78%→82.71%), with smaller gains in overall course mean (76.69→77.20) and pass rate (93.62%→94.90%). In short, all headline indicators moved favourably, the examination pass rate most.

Table 6. Comparison of B-2 and B-3 Academic Performance

Indicator	Coverage	B-2	B-3	Change
Weighted examination mean	51 classes	70.60	72.23	+1.63 points
Examination pass rate	51 classes	76.78%	82.71%	+5.93 pp
Weighted overall course mean	51 classes	76.69	77.20	+0.50 points
Overall course pass rate	51 classes	93.62%	94.90%	+1.28 pp

Note: pp = percentage points. Changes were calculated before rounding aggregate results; displayed rounded means may therefore differ by 0.01 from direct subtraction. Pass rates were recalculated from participant and failure counts. Source: institutional course statistics for the two semesters.

Examination failures fell from 499 to 346 and overall course failures from 137 to 102; pass rates were 76.78%→82.71% (exams) and 93.62%→94.90% (overall). The drop in failures, not just the rise in averages, is what makes the shift meaningful. Class by class, 33 of 51 (67.35%) raised their examination mean and 31 (60.78%) their overall mean; a fixed B-2 weighting gave 70.60→72.47 (exams) and 76.69→77.32 (overall). The upward move survived a different weighting, so it is unlikely an aggregation artefact. Among process components, individual assignments rose most (79.24→81.96, +2.72); classroom performance slipped (87.35→83.49), as did stage tests (84.34→83.63) and group tasks (79.11→78.71). Gains clustered where independent, criterion-referenced work was assessed, while teacher-scored participation eased back.

Across the six subsections the results tell a consistent story in two registers: the questionnaire shows students who knew the objectives, leant on AI for the basics and revision, and felt they improved while aware of its limits; the class records corroborate from outside, with the examination pass rate and individual-assignment scores carrying the clearest signal. Neither proves causation, but the convergence gives the early outcomes their weight.

5. Discussion

The questionnaire and course records describe the early outcomes complementarily. Students rated the clear requirements, AI support, engagement, and gains positively, and the cohort indicators moved the same way. The 5.93-pp rise

in the examination pass rate is the clearest signal that more students cleared the bar; the already-high overall pass rate crept up too. Together the two sources give a grounded account of what this design looked like in one setting—an implementation described, not a hypothesis confirmed.

5.1 Explicit Outcomes as a Reference for Learning

The high share who could state the objectives and criteria suggests outcome orientation was something students noticed, not just syllabus text. Explicit expectations let students tell what matters and use feedback against a known target, matching the constructive-alignment claim that coherence among outcomes, activities, and assessment makes a course legible [2].

A presentation task makes this concrete: require English sources, an organised explanation, and room questions; score content, structure, language, and interaction; have each student draft an independent outline, get AI suggestions, and state in a sentence which they took and why. Such arrangements turn positive ratings into something repeatable.

5.2 AI Feedback and Autonomous Revision

The warm reception for comprehension support, spotting expression problems, and revising shows feedback and revision deserve a central, not marginal, place. Students also reported finishing tasks on their own initiative and planning later study—behaviours suggesting they valued the process for managing learning, not shortcutting answers. The literature is unsettled: Escalante, Pack, and Barrett [7] found AI writing feedback performed about as well as a human tutor, while Barrot [8] warns that over-reliance can dull the critical and creative thinking writing should train.

A workable sequence: produce first alone, run an AI check, discuss with peers, then bring it to the teacher. This preserves the student's voice while keeping help timely; the aim is to help students understand and fix work against the criteria, not receive a correction.

5.3 Multiple Feedback Sources and Responsible Use

That students checked AI's facts and kept their views shows they know judgement is theirs. The worries about dependence and unreliable feedback show where instruction must still work; Hockly [9] argues such risks—inaccuracy, bias, over-reliance—should be named openly. A teacher can model verification on a real task and ask students to justify accepted or rejected suggestions, so teacher comments, peer review, and AI feedback pull together—as Kohnke et al. [4] urge through explicit digital literacy.

Assessment can weight product, revision process, and live performance together: draft independently, revise with assistance, then do a transfer task alone. This gives feedback access and a chance to show independent performance, embedding UNESCO's human-centred caution [6] in routine practice.

5.4 Sustaining Outcomes Through Course Evaluation

The rise in examination and individual-assignment scores anchors the positive survey responses. Continued evaluation can track these indicators and add unit performance tasks, revision portfolios, and feedback records, letting teachers adjust difficulty and support to keep goals, practice, and assessment connected.

A useful next step would follow individual performance with common standards and comparable tasks, extending this descriptive evidence to show whether the features persist across stages and learner groups.

6. Conclusion

Based on 329 usable questionnaires and 51 shared teaching classes in the 2024-entry cohort, the AI-enhanced, outcome-based course showed encouraging early outcomes. Students reported clear objectives and assessment, valued AI explanations and revision help, and felt gains in engagement and the language; the weighted examination mean rose 1.63 points, the pass rate 5.93 pp, and overall indicators edged up. The evidence supports carrying forward explicit outcomes, task-based practice, and timely feedback in this context.

The study has limits. It rests on single-institution questionnaire and class-level data; survey responses and grades were not matched student by student, we did not confirm equivalent exams or scoring across stages, and a few AI-use items showed inconsistent patterns. The findings describe early outcomes, not causes. Individual longitudinal records and common assessment tasks would show whether what we saw lasts and travels.

 Copyright @ STEMM Institute Press

References

[1] Spady, W. G. (1994). Outcome-Based Education: Critical Issues and Answers. Arlington, VA: American Association of School Administrators.

[2] Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364.

[3] British Council. (2024). Artificial intelligence and English language teaching: Preparing for the future. London: British Council. https://teachingenglish.britishcouncil.org/artificial-intelligence-and-english-language-teaching-research-british-council

[4] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550.

[5] Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

[6] Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. Paris: UNESCO. https://www.unesco.org/en/artificial-intelligence/education

[7] Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. International Journal of Educational Technology in Higher Education, 20(1), Article 57.

[8] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, Article 100745.

[9] Hockly, N. (2023). Artificial Intelligence in English language teaching: The good, the bad and the ugly. RELC Journal, 54(2), 445–451.

Copyright @ STEMM Institute Press