Project-Driven Learning and Generative AI Collaboration in Web Development Education: A Reform Study and Practice

Qiuxue Ouyang, Jilong Wu, Yajie Yu, Zhongda Yu, Weiyi Liu, Jinxiang Feng*

Qingdao City University, Qingdao 266000, China

jinxiang.feng@qdc.edu.cn

Abstract. This study addresses the disconnect between theory and practice, and the absence of generative AI tools, in conventional Web development courses (HTML5+CSS3). A hybrid instructional reform model integrating Project-Based Learning (PBL) with generative AI assistance (PBL-GenAI) was proposed and empirically validated. A sample of 480 students from seven higher vocational and applied undergraduate institutions participated in a multi-group quasi-experimental design (traditional instruction / PBL-only / PBL-GenAI). Reform effectiveness was assessed using multi-dimensional scales with rigorously verified reliability and validity. Results indicate that the PBL-GenAI group outperformed the traditional group by 15.3% on the composite weighted score and 21.6% on project deliverable quality; classroom engagement and learning satisfaction improved by 29.8 and 31.2 percentage points, respectively (all $p<0.001$). A three-stage AI literacy framework—Prompt Engineering, Critical Code Review, and Human-AI Collaboration Protocol—was concurrently developed to mitigate the risk of cognitive offloading.

Keywords: web development; HTML5+CSS3; project-based learning; generative AI; GitHub Copilot; blended learning; AI literacy

1 Introduction

PBL has accumulated substantial empirical support in engineering education [1,2], yet most studies employ small samples (<100 participants) [3,4,5] and provide insufficient validity reporting. Regarding generative AI in programming education, Prather et al. [6] found that GitHub Copilot benefits learners with stronger metacognitive skills while leaving weaker learners prone to debugging (dead-end loops). ICER 2025 further revealed that Copilot users spent 10.63% less time writing code manually, yet some students admitted they "did not understand why the AI suggestion worked," exposing a fundamental tension between efficiency gains and depth of understanding [7,12]. The OECD Digital Education Outlook 2026 explicitly notes that the learning benefits of generative AI are only sustained when its use is guided by clear pedagogical intent [8]—providing the core theoretical basis for this study's purposive AI collaboration

C. V. Garces et al. (eds.), *Proceedings of the 6th International Conference on Internet, Education and Information Technology (IEIT 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 53,
https://doi.org/10.2991/978-94-6239-733-0_37

framework. Three research gaps remain: the absence of an integrated PBL-GenAI framework; no empirical study targeting Web development courses specifically; and pervasive under-reporting of measurement instrument validity.

2 The PBL-GenAI Collaborative Pedagogical Reform Framework

2.1 Overall Framework Design

The reform model adopts OBE (Outcomes-Based Education) as the overarching logic, PBL as the core instructional strategy, generative AI as a cognitive scaffold, blended learning as the delivery mode, and multi-dimensional formative assessment as the quality assurance mechanism—forming an integrated five-component framework. AI tools are systematically embedded at every critical stage of PBL project workflows, while the three-stage AI literacy framework simultaneously guards against over-reliance. The overall framework is presented in Figure 1.

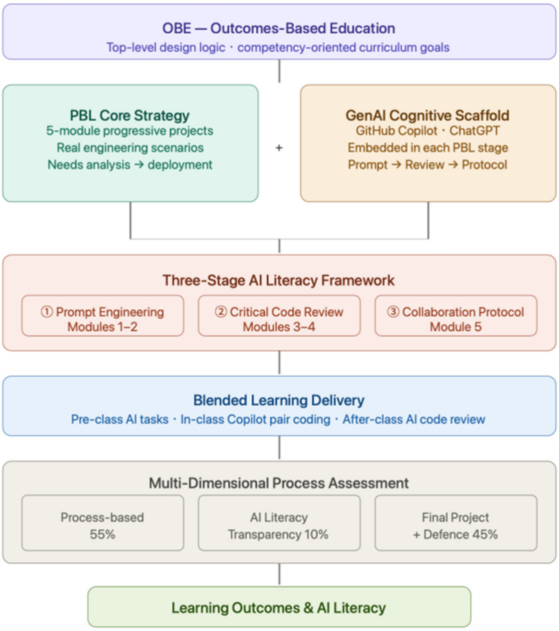

Fig. 1. Overall Framework Diagram of the PBL-GenAI Collaborative Teaching Reform Plan.

2.2 Three-Stage AI Literacy Framework

To mitigate cognitive offloading, a three-stage framework spans the course: Stage 1—Prompt Engineering (Modules 1–2): students independently formulate requirements and convert them into effective prompts, then evaluate AI outputs against expectations to enforce active cognition; Stage 2—Critical Code Review (Modules 3–4): instructors embed deliberate errors (e.g., Flexbox shorthand misuse, incorrect media query breakpoints)[9] in AI-generated code, which students must identify and justify, turning AI fallibility into deep learning; Stage 3—Human–AI Collaboration Protocol (Module 5): submissions must be line-by-line explainable, with Git commits documenting AI-generated, human-reviewed, and human-modified code proportions, ensuring transparent and assessable AI usage.

2.3 Three-Phase Blended Instructional Model

Pre-class phase (online, ~70 min): an "AI Exploration Task" is embedded in the structured preview package, requiring students to pose three questions on core content to ChatGPT and log both accurate and inaccurate responses. In-class phase (face-to-face, 90–100 min): a "Copilot Pair Programming" segment is introduced, modelling the workflow of independent reasoning → writing comments/pseudocode → activating Copilot suggestions → comparative analysis → adoption decision. Post-class phase: Gitee Pages combined with an AI code-review bot provides automated feedback, while final grading is determined by human instructor evaluation.

3 Research Design

3.1 Experimental Design and Sample

A multi-group quasi-experimental design was employed across 12 natural classes at seven institutions (480 students): Control Group C (n=160, traditional lecture-based instruction), Experimental Group E-PBL (n=160, PBL blended learning without AI tools), and Experimental Group E-GenAI (n=160, PBL-GenAI collaborative instruction). A pre-test administered at the start of semester (25 items; S-CVI/Ave=0.89) confirmed no significant baseline differences across groups ($F(2,477)=1.24$, $p=0.29$). Institution and class were entered as random effects in a hierarchical linear model (HLM) to control for inter-institutional variation.

3.2 Measurement Instruments and Psychometric Properties

Four measurement instruments, each pre-validated with samples of 200 or more, were used in this study (see Table 1). All scales employed a 5-point Likert format; item-level content validity indices (I-CVI) were ≥0.80; discriminant validity was confirmed via the Fornell-Larcker criterion (AVE >0.50 for all constructs).

Table 1. Psychometric Properties of Measurement Instruments

Scale Name	Items / Factors	α	ω	CFA Fit	Source
Classroom Engagement Scale	18 items, 3 factors	0.87	0.88	CFI=0.93, RMSEA=0.056	Fredricks et al., adapted
Learning Satisfaction Scale	20 items, 4 factors	0.89	0.91	CFI=0.91, RMSEA=0.061	Kuo et al., adapted

AI Tool Use Perception Scale	15 items, 3 factors	0.91	0.92	CFI=0.94, RMSEA=0.049	Nemt-allah et al., 2024 [10]
Technical Transfer Self-Rating Scale	12 items, 3 factors	0.85	0.86	CFI=0.90, RMSEA=0.064	Developed for this study
Scale Name	Items / Factors	α	ω	CFA Fit	Source

3.3 Data Analysis Strategy

Data were collected at three time points: start of semester, Week 10, and end of semester. Quantitative data were analysed using SPSS 27 and R 4.3, applying HLM (to account for nested data structures) and ANCOVA (with pre-test scores as the covariate); effect sizes were dual-reported as partial η^2 and Cohen's d, with Bonferroni correction for multiple comparisons. Qualitative data were drawn from semi-structured interviews with 40 students from the E-GenAI group; two independent coders achieved Cohen's κ=0.84.Qualitative data were coded using Braun and Clarke's thematic analysis framework [11].

4 Results

4.1 Learning Outcomes Across Three Groups

HLM analysis confirmed statistically significant differences among all three groups on every core outcome measure after controlling for nesting effects (see Table 2). Following Bonferroni correction, E-GenAI outperformed Group C on all measures ($p<0.001$; Cohen's d ranging 0.87–1.52, large effect sizes). Compared with E-PBL, E-GenAI showed significant advantages on project deliverable quality (d=0.37), learning satisfaction (d=0.89), and technical transfer self-rating (d=0.82) ($p<0.05$), demonstrating an incremental benefit attributable to AI integration beyond PBL alone.

Table 2. Learning Outcomes Across Three Groups (ANCOVA-adjusted means ± SE, controlling for pre-test)

Outcome Measure	Group C (Traditional)	Group E-PBL	Group E-GenAI	F	p	Partial η²
Final theory test (/100)	73.8±1.2	80.6±1.1	84.9±1.0	47.23	<0.001	0.165
Project deliverable quality (/100)	70.5±1.3	82.1±1.1	85.7±1.0	69.14	<0.001	0.225
Composite weighted score (/100)	72.4±1.2	81.5±1.0	83.5±0.9	55.38	<0.001	0.188
Classroom engagement (scale/100)	61.2±1.4	80.8±1.2	91.0±1.0	124.70	<0.001	0.344
Learning satisfaction (scale/100)	65.4±1.5	83.7±1.1	96.6±0.9	148.52	<0.001	0.384
Technical transfer self-rating (scale/100)	62.1±1.3	77.4±1.2	88.9±1.0	108.43	<0.001	0.312

4.2 Differential Effects of AI Tool Use

Students in E-GenAI were divided by AI usage transparency score into a Compliant Use group (≥70 points, n=97) and a Low-Transparency group (<70 points, n=63). The Compliant Use group achieved significantly higher end-of-semester scores (M=87.3±7.1 vs. M=79.6±9.8), $t(158)=5.42$, $p<0.001$, $d=0.89$, validating the predictive validity of the transparency dimension within the three-stage framework. Qualitative analysis revealed that approximately 22% of students exhibited a "copy-paste-submit-unable to explain" pattern at the start of the course; following instructor intervention, this proportion fell to 7% by Week 8. Among interviewees, 88% identified the Code Defence mechanism as the primary driver of their motivation to genuinely understand AI-generated output.

4.3 Supporting Effects for Lower-Achieving Students

Tracking students in the bottom 25% on the pre-test (n=120, 40 per group), E-GenAI students achieved a significantly higher composite score (76.2±9.3) than E-PBL (68.4±10.1) and Group C (61.3±11.7), $F(2,117)=18.74$, $p<0.001$, partial $\eta^2=0.242$, confirming that generative AI provides a meaningful compensatory support effect for lower-achieving learners when constrained by a structured literacy framework.Regarding the generalizability of the three-stage AI literacy framework: while it demonstrated robust effectiveness in this Web development context (HTML+CSS), its applicability to other computing disciplines warrants consideration. A key characteristic of front-end Web development is the visual immediacy of feedback—students can instantly observe whether a CSS layout renders correctly, making AI-generated errors in properties such as Flexbox alignment or media query breakpoints readily detectable through browser inspection. This tight output-verification loop may amplify the effectiveness of Stage 2 (Critical Code Review), because errors surface naturally during normal testing. By contrast, courses with less immediate feedback loops—such as data structures, operating systems, or theoretical algorithm design—may not afford learners the same low-cost error-detection opportunities. In those domains, AI-generated code may appear syntactically correct yet embody subtle logical flaws that manifest only under specific edge cases, potentially reducing students' motivation or ability to engage in genuine critical review. Future implementations in such courses should therefore supplement Stage 2 with explicit test-case design exercises and structured peer code inspection protocols, so that the absence of visual immediacy does not undermine the framework's core mechanism of turning AI fallibility into a deep learning opportunity.

5 Conclusion

The PBL-GenAI model improved the composite weighted score by 15.3% over traditional instruction and by an additional 2.5% over PBL alone. This moderate incremental gain suggests that, within an effective pedagogical framework, AI functions as a quality enhancer rather than a disruptive transformation—with its greatest added value manifest in learning satisfaction (+12.9 percentage points over E-PBL) and technical transfer self-rating (+11.5 percentage points), indicating that the primary contribution of AI collaboration lies in enriching the learning experience and deepening conceptual understanding. The significant association between AI usage transparency and academic outcomes (d=0.89), together with the behavioural change driven by the Code Defence mechanism, aligns closely with the core finding of the OECD Digital Education Outlook 2026 [6] and collectively validates the three-stage AI literacy framework.

Limitations: sample limited to domestic vocational/applied undergraduate contexts; transparency measured via Git logs; single-semester duration.

Future work: Three directions are prioritised. First, learning analytics will be integrated to enable finer-grained modelling of student–AI interaction patterns; specifically, we plan to apply process-mining techniques to Git commit histories and Copilot event logs to identify productive versus superficial AI engagement trajectories at the individual learner level. Second, the PBL-GenAI model will be extended to component-based front-end frameworks (Vue 3 / React) and full-stack courses (Node.js + database integration), where project complexity and multi-file architecture introduce qualitatively different cognitive demands and AI-assistance dynamics that may require revisions to the three-stage framework. Third, longitudinal tracking across at least two academic years will be conducted to examine whether AI literacy gains observed within a single semester persist into subsequent courses and transfer to workplace settings.

References

1. Yuan, X., Wan, J., An, D., Lu, J., Yuan, P.: Multi-method integrated experimental teaching reform of a programming course based on the OBE-CDIO model. Scientific Reports 14, 16623 (2024). https://doi.org/10.1038/s41598-024-67667-6
2. Hu, Z., Chang, Y.F., Ho, M.K.: A comparative study on the effectiveness of blended learning, physical learning, and online learning in functional skills training among higher vocational education. STEM Education 4(3), 247–262 (2024). https://doi.org/10.3934/steme.2024015
3. Xu, Z., Ge, L., He, W., et al.: Developing a BOPPPS model combined with the OBE concept to improve teaching outcomes of higher education. Humanities and Social Sciences Communications 11, 1001 (2024). https://doi.org/10.1057/s41599-024-03519-y
4. Boguslawski, S., Deer, R., Dawson, M.G.: Programming education and learner motivation in the age of generative AI: student and educator perspectives. Information and Learning Sciences 126(1/2), 91–109 (2025). https://doi.org/10.1108/ILS-10-2023-0163
5. Guibord, G., et al.: A systematic review of research on teaching web design to secondary school students through HTML and CSS. Discover Education 4(1), 357 (2025). https://doi.org/10.1007/s44217-025-01079-0
6. Finnie-Ansley, J., Denny, P., Becker, B.A., Luxton-Reilly, A., Prather, J.: The robots are coming: Exploring the implications of OpenAI Codex on introductory programming. In: Proc. 24th Australasian Computing Education Conference (ACE '22), pp. 10–19. ACM, New York (2022). https://doi.org/10.1145/3511861.3511863
7. Shihab, M.I.H., Hundhausen, C., Tariq, A., Haque, S., Qiao, Y., Wise, B.M.: The effects of GitHub Copilot on computing students' programming effectiveness, efficiency, and processes in brownfield coding tasks. In: Proc. ACM Conference on International Computing

Education Research V.1 (ICER '25), pp. 1–14. ACM, New York (2025). https://doi.org/10.1145/3702652.3744219
8. OECD: OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education. OECD Publishing, Paris (2026). https://doi.org/10.1787/062a7394-en
9. State of CSS 2024 Survey Results. https://2024.stateofcss.com (2025)
10. Nemt-allah, M., Khalifa, W., Badawy, M., Elbably, Y., Ibrahim, A.: Validating the ChatGPT Usage Scale: psychometric properties and factor structures among postgraduate students. BMC Psychology 12, 497 (2024). https://doi.org/10.1186/s40359-024-01983-4
11. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology 3(2), 77–101 (2006). https://doi.org/10.1191/1478088706qp063oa
12. Peng, S., Kalliamvakou, E., Cihon, P., Demirer, M.: The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:2302.06590 (2023). https://arxiv.org/abs/2302.06590