

This work is licensed
under a Creative
Commons Attribution
4.0 International
License.

ChatGPT's effects on EFL learners' autonomy and self-efficacy in self-regulated learning contexts

 **Quang Nhat Nguyen**
Saigon International University, Vietnam
nguyennhatquang@siu.edu.vn

 **Dung Thi Hue Doan**
Saigon International University, Vietnam
doanthihuedung@siu.edu.vn

This study compares the effects of ChatGPT-facilitated and human teacher-facilitated project-based learning (PBL) on learner autonomy and self-efficacy among Vietnamese EFL students in a language training center. Using an explanatory sequential mixed-methods design, fifty participants were randomly assigned to either a ChatGPT-facilitated group or a teacher-facilitated group for an eight-week project-based intervention. Learner autonomy and self-efficacy were measured before and after the intervention. Quantitative data were analyzed using independent t-tests and one-way ANCOVA, while qualitative data from focus group interviews were coded and thematically analyzed. The results indicated that the ChatGPT group achieved significantly greater gains in learner autonomy, whereas both groups improved equally in self-efficacy. Qualitative analysis revealed that AI facilitation transformed learner perception of autonomy and self-efficacy and encouraged group initiative and experimentation, while teacher facilitation reinforced established routines and delegated authority. These findings demonstrate the potential of AI-driven facilitation to promote learner autonomy and suggest that blended models combining AI and human strengths may optimize language learning outcomes.

Keywords: project-based learning, ChatGPT, artificial intelligence, learner autonomy, self-efficacy

Introduction

The rapid integration of artificial intelligence (AI) tools, particularly sophisticated language models like ChatGPT, into educational settings has introduced

transformative opportunities for learner-centered instruction and innovative pedagogical practices. Concurrently, Project-Based Learning (PBL), a constructivist pedagogical approach characterized by collaborative and authentic learning activities, has become widely adopted due to its proven efficacy in enhancing learner engagement, critical thinking skills, and language competencies (Bell, 2010; Thomas, 2000). Extensive empirical research supports the educational benefits of PBL. Bell (2010) emphasizes PBL's ability to stimulate learners' intrinsic motivation, facilitate collaboration, and promote active learning. Thomas (2000) similarly underscores its role in developing higher-order cognitive skills such as critical thinking and problem-solving.

Although research on project-based learning and learner autonomy has flourished in developed countries, the adoption and impact of such pedagogical innovations in Asian educational systems, especially in Southeast Asia, remain critically underexamined. In much of Asia, English language education continues to be shaped by examination-driven curricula, centralized policies, and enduring cultural norms prioritizing teacher authority and collective harmony, often at the expense of individual agency and learner voice (Chan, 2001; Littlewood, 1999). This reality creates a persistent tension between progressive, learner-centered reforms and entrenched practices that discourage autonomy and risk-taking.

Recent reform initiatives in Vietnam and other Asian countries have signaled an official shift toward communicative competence, task-based learning, project-based activities, and digital integration. However, these reforms face significant implementation challenges, including long-lasting mental inertia, insufficient teacher training, resource disparities, and variable digital literacies across urban and rural regions (Duc et al., 2022, Vu & Nguyen, 2021). The introduction of generative AI, such as ChatGPT, offers new potential for individualized learning and immediate feedback (Zhang & Wang, 2025). A large body of studies highlights how the integration of AI, particularly ChatGPT, has enhanced educational experiences through personalized, instantaneous feedback and advanced prompting techniques, significantly influencing learners' self-regulation skills and overall autonomy (Celik et al., 2025; Xu et al., 2024). However, integrating AI into language education also raises questions about equity, authenticity, and the further marginalization of students who lack the requisite digital skills or critical thinking habits fostered by more dialogic, human-centered facilitation.

While previous research has extensively evaluated the effectiveness of PBL and AI-facilitated instruction through observable outcomes such as language performance, assessment scores, and task completion (Zhu & Wang, 2025), these outcome-driven metrics offer only a partial picture of the learning experience. In AI-mediated, project-based environments, internal constructs such as learner motivation, emotional safety, autonomy, and self-efficacy become beneficial and essential, as affective dimensions fundamentally shape how learners engage, persist, and make meaning in complex, collaborative tasks.

Moreover, research on learning outcomes has reached a degree of saturation, especially in cognitive domains. How facilitation mode, AI or human,

modulates learners' perceptions of agency, control, and emotional investment in the learning process remains underexplored. Without addressing these psychological drivers, even the most impressive outcome gains may lack long-term impact or fail to empower students beyond the instructional setting. By prioritizing autonomy and self-efficacy, this study contributes to a growing recognition that deep, sustainable learning depends not only on what learners achieve but also on how they experience the process, a perspective aligned with ecological and humanistic approaches to language learning (Benson, 2011; van Lier, 2004).

Despite comprehensive research on the independent educational contributions of PBL and AI-driven instruction, critical gaps remain concerning comparative analyses of AI versus human facilitation within PBL frameworks. Specifically, little empirical attention has been paid to understanding how these distinct modes of facilitation differentially impact essential psychological constructs like learner autonomy and self-efficacy, particularly in lower-intermediate proficiency contexts. Responding to this need, the present study systematically compares the impact of ChatGPT-facilitated versus human teacher-facilitated PBL among Vietnamese EFL learners at A2-B1 proficiency levels. The study aims to address the following research questions:

- Are there significant differences in learner autonomy between learners facilitated by ChatGPT and those facilitated by human teachers?
- Are there significant differences in general self-efficacy between these groups?
- How do learners perceive and articulate the influences of ChatGPT versus human teacher facilitation on their autonomy and self-efficacy?

Through these inquiries, this research aims to contribute meaningfully to theoretical understanding and practical applications regarding effective AI integration in language education.

Literature review

Project-based learning (PBL)

Project-based learning (PBL) is extensively conceptualized as an instructional methodology grounded in constructivist educational theory, emphasizing student-centered inquiry, collaboration, and authentic learning experiences. Thomas (2000) states that PBL involves students in sustained, collaborative tasks that address complex, real-world challenges, resulting in tangible, meaningful products. Bell (2010) further elaborates that PBL fosters deeper cognitive engagement by promoting active participation, interaction, and critical reflection throughout the learning process. Operationally, this study defines PBL as an instructional approach where students collaboratively engage in authentic projects that culminate in concrete outcomes and require application, analysis, and knowledge synthesis.

Learner autonomy

Learner autonomy has been widely examined within educational and applied linguistics frameworks, often defined as learners' capability to direct their educational journeys independently. Holec (1981) initially characterized learner autonomy as learners' ability to take charge of their learning processes, including setting clear objectives, selecting suitable learning strategies, and effectively assessing their progress and outcomes. Benson (2011) expands this definition by highlighting the significance of learners' willingness and ability to make independent choices regarding their learning experiences and emphasizing the necessity of critical reflection and self-assessment skills. Given the general definitions of learner autonomy, this concept in the current study is defined as the learner's perceived capacity to set personal learning goals, plan learning activities, implement learning strategies, monitor progress, and evaluate outcomes in foreign language learning. This operationalization is aligned with the multidimensional construct measured by the Learner Autonomy Perception Questionnaire (LAPQ) developed and validated by Nguyen and Habók (2021), which encompasses desire, motivation, freedom, metacognition in language learning, and beliefs about teachers' role. However, as the facilitator in this study is not merely a teacher, the authors will change the final dimension to beliefs about the facilitator's role.

Self-efficacy

Self-efficacy, conceptualized within Bandura's (1977) social cognitive theory, refers to individuals' beliefs about their ability to execute necessary actions to achieve desired outcomes effectively. Bandura emphasizes that self-efficacy significantly impacts motivation, cognitive processes, affective states, and behavioral actions. High self-efficacy fosters persistence, effort, and resilience in facing difficulties and challenges. Schunk and Pajares (2002) further assert the importance of self-efficacy within educational contexts, noting its direct influence on learners' goal setting, motivation, persistence, and academic performance. Self-efficacy in this research refers to learners' confidence in their ability to effectively carry out specific language learning tasks (listening, speaking, reading, writing, and class participation) in English. Operationally, this study defines self-efficacy as learners' self-reported confidence in their capacity to engage in and complete tasks successfully within the PBL activities, influenced by their experiences, feedback, and perceived mastery during the project implementation. This definition corresponds to the construct measured by the Foreign Language Learning Self-Efficacy Questionnaire (FL-LSEQ) (Kutuk et al., 2023), which assesses students' self-appraisals across core domains of language performance and classroom engagement as they receive and produce the language.

Recent research classifies self-regulated learning (SRL) models into several types, with most frameworks recognizing at least three main paradigms. First, teacher-regulated models emphasize direct instruction and frequent scaffolding; here, the teacher guides goal-setting, planning, and monitoring (Panadero, 2017). Second, co-regulated or shared-regulation models balance agency between teachers, peers, and students, allowing guided autonomy within structured parameters. Third, self-regulated models place learners at the center, requiring them to initiate, manage, and evaluate their learning cycles (Zimmerman & Schunk, 2011). Within self-regulation, various subtypes exist, ranging from partially scaffolded approaches to those advocating high autonomy but with strategic checkpoints (Ifenthaler & Schumacher, 2016).

Globally, most empirical SRL research and classroom practice remain in the domain of guided or “supported” self-regulation, where teachers design scaffolds and technologies mediate learning, but learners’ freedom is still bounded by preset objectives, timelines, or rubrics (Holstein et al., 2020). Even project-based learning models that claim to be “student-driven” rarely allow unrestricted agency; instead, freedom is often nominal and circumscribed by frequent teacher intervention or rigid project milestones (Bell, 2010; Panadero, 2017).

The notion of “total freedom”, where students are granted authentic, unrestricted control over project choice, timeline, collaboration, strategies, and even the criteria for success, has been largely absent in SRL literature and large-scale educational practice. Leading systematic reviews acknowledge that, despite frequent references to “learner autonomy,” almost all so-called “autonomous” models implement extensive monitoring, require predefined project rubrics, or reserve final authority for teachers or algorithms. There are conceptual arguments for radical freedom (Biesta, 2020), but published empirical studies with “total freedom” in real classrooms are exceedingly rare. The main reasons cited are concerns over student floundering, institutional accountability, and difficulty measuring outcomes (Panadero, 2017; Thomas et al., 2019).

This study represents one of the first empirical implementations of total freedom in project-based, self-regulated learning for EFL students. Here, students are not just given a choice within fixed constraints but are trusted to define every aspect of their learning project, including topic, process, timeline, resource use, and self-assessment criteria. Teachers step back fully, intervening only for ethics or safety, and no artificial “scaffolds” are imposed to simulate agency. The approach is documented with transparent criteria and outcome analysis, making it a significant contribution and a benchmark for truly learner-centered education (Immordino-Yang et al., 2018). This way, the research breaks from the global norm, offering the field a rare, authentic model of total autonomy in higher education.

Effects of AI and teacher facilitation on self-efficacy

The literature documents robust effects of project-based learning on self-efficacy, especially when tasks are authentic, collaborative, and reflective (Bell, 2010; Thomas, 2000; Yildiz, 2024). Bandura's social cognitive theory emphasizes the importance of supportive feedback, observational learning, and mastery experiences in building self-efficacy. Traditionally, teacher-led PBL environments have enhanced students' confidence, resilience, and willingness to take risks. Recent investigations into AI-based facilitation report similarly promising results, citing increases in learner self-confidence, willingness to participate, and resilience, particularly when AI provides individualized and timely feedback. Notably, Zhang et al.(2025) used surveys and classroom analytics to explore the relationship between AI literacy and willingness to communicate among Chinese university EFL students. They found that students with higher AI self-efficacy and lower anxiety benefited most from AI interventions, while those lacking digital skills remained marginalized. No studies have directly compared ChatGPT- and teacher-facilitated PBL in Vietnamese EFL settings, particularly among lower-intermediate learners or outside major urban centers. However, the existing research rarely compares these two facilitation modalities directly. Furthermore, comprehensive multidimensional measures of self-efficacy, encompassing academic and socio-emotional domains, are infrequently employed. This omission limits understanding of how different modes of facilitation might uniquely impact the full spectrum of learner self-efficacy.

Effects of AI and teacher facilitation on learner autonomy

Fostering learner autonomy is a key goal in both PBL and AI-enhanced classrooms. Bell (2010) and Thomas (2000) note that PBL, through authentic, student-centered projects, develops autonomy by encouraging planning, decision-making, and reflective learning. Research from East Asia raises important challenges. Littlewood (1999) and Chan (2001), using survey-based designs with university students in Hong Kong and other East Asian regions, found that strong cultural norms supporting teacher authority and exam success limited the development of genuine autonomy in PBL environments.

An emerging body of research highlights that AI-driven tools can further support autonomy by promoting self-regulation, information-seeking, and independent inquiry. A study by Wu et al. (2025) convincingly demonstrates that an AI-supported multimodal learning environment significantly improves the critical thinking and self-regulation of primary school EFL students in China compared to traditional methods. While this research is valuable for its robust mixed-methods design and practical results, its focus on young learners highlights a significant gap when considering university students. These findings are not directly generalizable to higher education, where students face more complex academic tasks, require a more advanced level of self-regulation, and critically need the ability to evaluate AI-generated content to ensure academic integrity. Also, there is still a lack of research focusing directly on autonomy.

Therefore, further research is needed to determine the effectiveness of these AI models for the distinct cognitive and academic demands of students in higher education. The introduction of AI thus adds a nuanced layer, where immediate access to resources and responsive feedback may support new forms of autonomous learning. Nevertheless, prior studies focus on surface-level indicators of independence, often neglecting deeper forms of autonomy such as metacognitive reflection or epistemic control. Moreover, there is a limited exploration of these dynamics among lower-intermediate EFL learners, which stands to benefit most from scaffolded autonomy and self-directed learning strategies.

The effectiveness of project-based learning in fostering autonomy and self-efficacy is extensively documented, particularly within traditional teacher-led frameworks. However, as educational technologies evolve, there is a pressing need to address how alternative facilitation methods, such as AI-driven approaches, compare and interact with established pedagogical models. Previous research has not sufficiently examined the direct and potentially complementary effects of AI and human facilitation on multidimensional constructs of autonomy and self-efficacy, nor has it explored these relationships among lower-intermediate EFL learners who critically require such support. The introduction of AI is theorized to enhance specific aspects of autonomy, such as immediate feedback and independent inquiry, while teacher-led facilitation is expected to influence the socio-emotional and relational facets of self-efficacy more strongly. This study aims to empirically test and refine this framework to identify how alternative facilitation models can promote learner autonomy and self-efficacy in contemporary language education (see Figure 1).

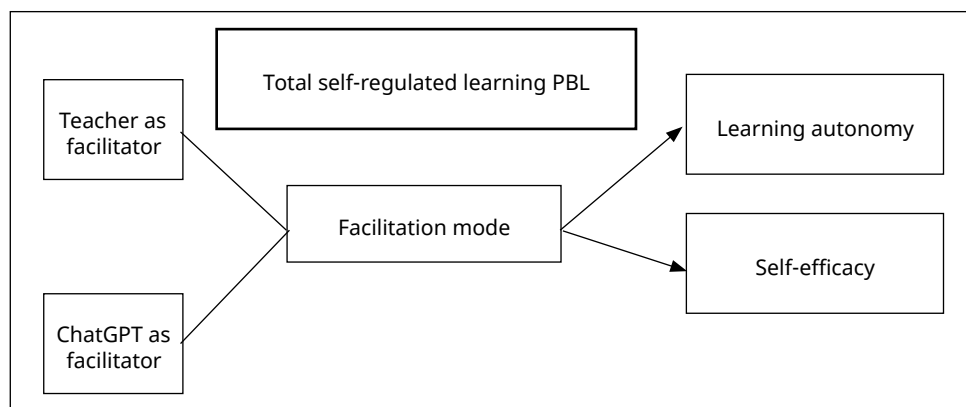


Figure 1. Conceptual framework

Method

Research design

This study employed an explanatory sequential mixed-methods design (Creswell & Clark, 2017), integrating quantitative and qualitative approaches to comprehensively understand how facilitation mode (ChatGPT versus human teacher)

affects learner autonomy and self-efficacy. The research was conducted in two distinct but connected phases. First, quantitative data were collected and analyzed using a quasi-experimental pretest-posttest control group design, which enabled statistical comparisons between groups. Second, qualitative data were gathered through semi-structured focus-group interviews to further explain, contextualize, and expand upon the quantitative findings. The explanatory sequential mixed-methods design is especially appropriate for educational intervention studies where understanding effect size and process is essential (Creswell & Clark, 2017; Ivankova et al., 2006).

Participants and sampling

The study occurred at a private language center in Ho Chi Minh City, Vietnam, where English is taught as a foreign language to help students meet university graduation requirements. This setting is typical of Southeast Asian EFL contexts, with exam-oriented curricula, high-stakes assessments, and a gradual move toward communicative, student-centered teaching (Dörnyei & Murphey, 2003; Little, 2007; van Lier, 2004). The center's research ethics committee granted ethical approval. Participation was voluntary, with informed consent, confidentiality, and the right to withdraw always ensured. The use of AI tools followed institutional policies and digital safety standards.

Fifty EFL learners aged 18 to 22 participated, with proficiency ranging from high-A2 to low-B1 on the CEFR. Two intact classes were randomly assigned to either a ChatGPT-facilitated or teacher-facilitated group ($n = 25$ per group). The gender balance was 60% female and 40% male, reflecting the broader student population preparing for English exit exams in southern Vietnam. Most students had only informal experience with ChatGPT or similar AI tools and had never participated in structured language learning using AI. Pre-study surveys confirmed no prior formal instruction with AI for language learning. Informed consent and the right to withdraw were guaranteed for all participants.

The study's use of intact classes was guided by organizational realities and foundational pedagogical considerations inherent to project-based learning. Authentic project-based learning depends on sustained collaboration, forming group relationships, and gradually developing collective norms. In actual educational environments, students are expected to negotiate roles, build trust, and engage with one another over time. Randomly assigning individuals to groups solely for methodological purposes may disrupt these processes and diminish the ecological validity of the intervention. Language learners are not experimental subjects in isolation; rather, their learning trajectories are shaped by the social dynamics and mutual support unique to each classroom community. By maintaining intact classes, this study seeks to preserve the natural conditions in which project-based learning unfolds and to produce findings that are meaningful for real-world educational practice. Although this approach imposes certain statistical limitations, it offers a more accurate representation of collaborative learning as it occurs in practice, and it provides insights that can inform both classroom instruction and future research. We

acknowledge the methodological limitation inherent in using intact classes for group assignments. To address this issue, additional measurement tools and monitoring procedures will be incorporated to better control for clustering effects and enhance the validity of future studies, which will be clearly elaborated in the methodology section.

For the qualitative phase, ten students from each group were purposively selected for focus group interviews to ensure a balance of age, gender, proficiency (CEFR A2–B1), and classroom roles (leadership, creative direction, technical support, research, peer mediation), supporting the interpretive power of the analysis (Creswell & Poth, 2018; Dörnyei & Murphey, 2003; Little, 2007; van Lier, 2004). Participant details are presented in Appendix 1. This purposive sampling reflects real project-based EFL dynamics and increases the credibility and transferability of the findings.

Project description and rationale

The project “A Day on English Television” required groups of five students to collaboratively design, script, and present an original English-language video inspired by authentic international TV shows. Over eight weeks, each group brainstormed several topics, selected suitable program formats such as talk shows, news, or documentaries, researched content, wrote detailed scripts, assigned production roles, rehearsed, and recorded a final video for assessment. Project requirements included the use of English in all group interactions, the development of a complete video product, demonstration of teamwork, and regular peer and facilitator feedback.

This project closely reflects the essential characteristics of project-based learning as outlined by Thomas (2000) and Blumenfeld et al. (1991). It centers on an open-ended, authentic task that requires sustained inquiry, collaboration, self-management, and application of knowledge to a real-world outcome. Students are engaged in extended, meaningful work and must solve complex problems, communicate ideas clearly, and adapt through multiple cycles of revision and feedback. The project structure allows for a robust comparison of facilitation models. In the ChatGPT-facilitated group, students received instruction and practice on how to use ChatGPT for language support, idea generation, and feedback, which encouraged greater learner autonomy and the exploration of multiple solutions. In the teacher-led group, the teacher provided formative feedback, modeled expert thinking, and guided collaborative processes. Both groups worked within a framework that emphasized student initiative and reflective practice, which are central to effective project-based learning.

Procedures

Ethical approval was obtained from the institutional review board. Written informed consent was collected from all participants before any research activity. Before the intervention, all participants completed validated pretest

measures of learner autonomy and self-efficacy under standardized conditions. Demographic data were collected at this stage to confirm baseline comparability.

Two intact classes were randomly assigned, by an independent administrator, to either the ChatGPT-facilitated or teacher-facilitated condition. Both groups attended a mandatory orientation session that explained the project objectives, assessment rubrics, group responsibilities, and ethical guidelines. The ChatGPT group received additional hands-on instruction on prompt engineering and safe use of ChatGPT; the teacher group was not exposed to AI tools.

The intervention lasted eight weeks, with each group attending one ninety-minute project-based learning session per week. Each session followed a fixed structure: (1) initial progress review, (2) collaborative planning, (3) project execution, and (4) structured reflection in every section. In the ChatGPT group, students interacted independently with ChatGPT for content generation, language support, and problem-solving; the teacher was present only for technical supervision and safeguarding ethical compliance. In the teacher-facilitated group, the teacher provided all instructional guidance on students' requests, formative feedback, and adaptive support, with no access to AI assistance. Both groups were subject to identical project requirements, timelines, rubrics, and resource access.

All project artifacts, group outputs, and weekly learning logs were systematically collected and archived throughout the intervention. Immediately after the eight-week intervention, participants in both groups completed posttest measures of autonomy and self-efficacy, under the same conditions as the pretest. A purposive subsample from each group participated in semi-structured focus group interviews, conducted in Vietnamese, to explore individual and group experiences in detail. Considering the nature of project-based learning, the interviewees were chosen to ensure a diversity of roles in projects rather than based on their autonomy and self-efficacy scores. Interview recordings were transcribed verbatim and backtranslated for analysis.

Quantitative data were analyzed using independent t-tests to ensure baseline levels of autonomy and self-efficacy, analysis of covariance (ANCOVA), with pretest scores as covariates, to test the effect of facilitation mode on posttest outcomes. All model assumptions were checked and satisfied. Qualitative data were analyzed using thematic analysis, with coding procedures independently verified by two researchers to ensure consistency and reliability until reaching the saturation point.

Instruments

Learner Autonomy Perception Questionnaire (LAPQ): Learner autonomy was measured using the validated Learner Autonomy Perception Questionnaire (LAPQ) developed and standardized by Nguyen and A. Habók (2021). The LAPQ has been specifically validated for EFL learners in Asian contexts, including Vietnam, demonstrating high internal consistency (Cronbach's $\alpha > .80$) and strong construct validity (Nguyen & Habók, 2021). Some wording of this scale has been updated to meet the nature of this study, where the facilitator is not

only a teacher but also ChatGPT. The instrument consists of 40 items scored on a 5-point Likert scale, covering five dimensions: metacognitive skills (15 items), beliefs about the facilitator's role (8 items), motivation and desire (5 items), freedom and opportunity (7 items), and metacognitive knowledge (5 items). Item wording was revised to reflect both teacher and AI facilitation and to align with the project-based EFL learning context (see Appendix 2).

Foreign Language Learning Self-Efficacy Questionnaire (FL-LSEQ): Self-efficacy was assessed using the Foreign Language Learning Self-Efficacy Questionnaire (FL-LSEQ), recently developed and validated by Kutuk et al. (2023). The FL-LSEQ contains 11 items rated on a 5-point Likert scale, evaluating students' confidence in various aspects of foreign language learning, including receptive (5 items) and productive self-efficacy (6 items). All items in the self-efficacy questionnaire were carefully adapted from the original scale (Kutuk et al., 2023) to ensure contextual relevance for project-based English learning and compatibility with both ChatGPT-facilitated and human teacher-facilitated environments. The wording of each item was modified to explicitly reference project tasks, project-related communication, and facilitator roles (teacher or AI), thus ensuring that each statement reflected the authentic communicative and academic demands of the PBL context. These adaptations preserve the construct validity of the scale while enabling meaningful comparison across experimental conditions (see Appendix 3).

It is important to note that adaptation followed the recommended principles of scale contextualization. The revised instruments were reviewed for face validity by two EFL experts familiar with both project-based and AI-assisted language learning. Cronbach's alpha was calculated for each scale based on the study sample, with results above .80 for both the LAPQ and FL-LSEQ, indicating acceptable reliability in this context. However, it is noteworthy that the tests are now adapted in this study; further validation using CFA and EFA could be conducted to increase the reliability and validity for larger-scale research.

Semi-structured focus group interviews: To triangulate and enrich quantitative findings, semi-structured focus-group interviews were conducted with 20 participants (10 per group) after the intervention. The interview protocol explored learners' experiences with the PBL process, the perceived influence of the facilitation mode on their autonomy and self-efficacy, and the challenges or benefits encountered. Rather than focusing on specific dimensions of autonomy and efficacy, the researchers created general questions that provide learners with more opportunities to express their general learning experiences (see Appendix 4).

Data collection and analysis

Pretest and posttest data from LAPQ and GSE were coded and entered into SPSS version 25 for statistical analysis. All quantitative analyses, descriptive statistics, independent samples t-tests, and one-way analysis of covariance (ANCOVA) were conducted using SPSS.

To address the methodological limitations associated with the use of intact

classes for group assignment, several additional procedures were implemented to monitor and control for potential clustering effects. First, baseline equivalence between groups was established through independent samples t-tests on all key pretest variables, including learner autonomy and self-efficacy. Prior to conducting the t-tests, the normality assumption for each variable within both groups was examined using Q-Q plots and formal statistical tests. Second, the intraclass correlation coefficient (ICC) was calculated using the one-way ANOVA method because this approach directly partitions the total variance of the outcome variable into between-group (class-level) and within-group (individual-level) components, with the following formula of ICC (Howell, 2018):

$$\frac{MS_{bet\ S} - MS_{residual}}{MS_{bet\ S} + (k - 1)MS_{residual}}$$

By comparing the mean square values obtained from the ANOVA, the ICC quantifies the proportion of variance attributable to class-level clustering. This method is particularly suitable when evaluating clustering effects in studies with a small number of groups or when multilevel modeling is not feasible due to limited sample size at the group level. These measures were designed to enhance the internal validity of the study and to provide a more nuanced understanding of how intact class assignment may influence learning outcomes. Recommendations for future studies include the use of additional process controls and, where feasible, the collection of data across a greater number of classes to further strengthen the robustness of findings.

The facilitation group, ChatGPT or teacher-led, was specified as the fixed factor. Posttest scores were set as the dependent variable, and pretest scores were the covariate. ANCOVA was chosen to control baseline differences and reduce error variance, ensuring a precise estimate of the group effect. All model assumptions were tested and confirmed, including normality, homogeneity of variance, linearity between the covariate and outcome, and homogeneity of regression slopes. Effect sizes were reported using partial eta squared.

Qualitative data from the focus group interviews were transcribed verbatim and imported into NVivo version 14 for coding and thematic analysis. The initial coding framework was developed based on validated constructs of learner autonomy and self-efficacy and then refined to capture emergent themes. Coding was performed independently by two coders, a member of the research team and an external coder who was not involved in any other stage of the research; each analyzed all transcripts. Inter-coder reliability was assessed using Cohen's kappa. Any discrepancies were resolved through discussion until a full consensus was reached. This process ensured objectivity and minimized the potential for investigator bias in mixed-methods studies. All coding steps and analytic decisions were fully documented for transparency and replicability. We employed the concept of thematic saturation as a stopping criterion for focus group data collection and analysis. Saturation was defined as the point at which no new themes, subthemes, or significant insights

emerged from additional interviews, indicating that the sample was sufficient to capture the full range of relevant perspectives

Quantitative and qualitative findings were integrated during the interpretation phase, following the logic of explanatory sequential mixed methods. This integration provided a comprehensive account of how the facilitation mode influenced learner outcomes and ensured that quantitative results were corroborated and deepened by qualitative evidence.

Results and findings

Quantitative results

Pretest results. Prior to the t-test analysis, for autonomy, the Kolmogorov-Smirnov and Shapiro-Wilk tests yielded p-values of .200 and .903, respectively, indicating no significant departure from normality. For self-efficacy, the Kolmogorov-Smirnov test was significant ($p = .004$), while the Shapiro-Wilk test was not ($p = .156$); based on the sample size and Q-Q plot inspection, the distribution was deemed sufficiently normal to justify the use of parametric analyses. Before the intervention, independent samples t-tests showed no statistically significant differences between the ChatGPT-facilitated group and the teacher-led group on both learner autonomy and self-efficacy pretest scores (see Tables 1 and 2). Also, the normality assumption for each variable within both groups was confirmed using Q-Q plots and formal statistical tests. This indicates both groups began with comparable baseline levels in both constructs.

Table 1. Pretest score for self-efficacy

		Levene's test for equality of variances		t-test for equality of means					95% confidence interval of the difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	Lower	Upper
Mean	Equal variances assumed	.581	.450	.769	48	.446	.08000	.10400	-.12910	.28910
	Equal variances not assumed	.581	.450	.769	47.642	.446	.08000	.10400	-.12914	.28914

The preliminary analysis of the pretest scores of general self-efficacy indicated no significant baseline differences between the groups, $t(48) = 0.769$, $p = .446$. The assumption of homogeneity of variances was met for both pretest and posttest analyses, as assessed by Levene's test ($p > .05$).

Table 2. Pretest score for autonomy

		Levene's test for equality of variances		t-test for equality of means					95% confidence interval of the difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	Lower	Upper
Mean	Equal variances assumed	.010	.920	.625	48	.535	.05200	.08318	-.11525	.21925
	Equal variances not assumed	.010	.920	.625	47.773	.535	.05200	.08318	-.11527	.21927

An independent samples t-test was conducted to assess the impact of the intervention on participant autonomy. A preliminary analysis of the pretest scores confirmed that there were no significant baseline differences in autonomy between the experimental and control groups, $t(48) = 0.625$, $p = .535$. The assumption of homogeneity of variances was met for both pretest and post-test analyses (Levene's test $p > .05$).

Posttest results

Self-efficacy. Regarding the efficacy pretest, homogeneity of variances was confirmed using Levene's Test ($p = .450$), indicating that the assumption of equal variances was satisfied. The intraclass correlation coefficient (ICC) for efficacy pretest, calculated using the one-way ANOVA method ($MS_{\text{between}} = 0.080$; $MS_{\text{within}} = 0.135$; $k = 25$), was -0.017 , suggesting negligible clustering at the class level for this variable. Given that the intraclass correlation coefficient (ICC) for autonomy at pretest was negligible, and the assumption of equal variances was satisfied, it is appropriate to proceed with ANCOVA to assess the intervention effects while controlling baseline differences. The assumption of homogeneity of regression slopes was tested by including the interaction term between group and the covariate in the model. The interaction was not statistically significant, $F(2, 47) = 2.281$, $p = .113$, indicating that the relationship between the covariate and the dependent variable did not differ significantly across groups. Therefore, the assumption was met, and ANCOVA was considered appropriate for subsequent analyses.

One-way ANCOVA was then conducted to compare the ChatGPT group and the teacher-led group while statistically controlling pretest self-efficacy scores.

Table 3. Self-efficacy posttest descriptive statistics

Group	Mean	Std. deviation	N
Human teacher	4.0267	.81569	25
ChatGPT	4.2500	.74263	25
Total	4.1383	.78021	50

Table 4. Self-efficacy posttest one-way ACOVA analysis

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial Eta squared
Corrected model	.658 ^a	2	.329	.531	.592	.022
Intercept	18.835	1	18.835	30.348	.000	.392
Pretest	.035	1	.035	.056	.813	.001
group	.649	1	.649	1.045	.312	.022
Error	29.169	47	.621			
Total	886.118	50				
Corrected total	29.828	49				

a. R squared = .022 (Adjusted R squared = -.020)

An ANCOVA was conducted to examine whether facilitation mode (ChatGPT vs. teacher-led) had a significant effect on learner self-efficacy while controlling for pretest scores. Levene's test was non-significant ($p = .353$), confirming the homogeneity of variances. The pretest covariate was not significant, $F(1, 47) = .056$, $p = .813$. After controlling for pretest, there was no significant difference in posttest efficacy between groups, $F(1, 47) = 1.05$, $p = .312$, partial $\eta^2 = .022$ (See Table 3 and Table 4). Given the unexpectedly non-significant effect of the pretest covariate in the ANCOVA, further analysis was conducted to assess whether class-level clustering might have influenced posttest outcomes. Specifically, the intraclass correlation coefficient (ICC) for posttest self-efficacy was calculated using the one-way ANOVA method, with class as the grouping factor. This approach allowed us to quantify the proportion of variance in posttest scores attributable to class membership, thereby clarifying whether group-level effects contributed to the observed results.

Homogeneity of variances was confirmed using Levene's Test ($p = .707$), indicating that the assumption of equal variances was satisfied. The intraclass correlation coefficient (ICC) for self-efficacy posttest, calculated using the one-way ANOVA method ($MS_{\text{between}} = 0.742$; $MS_{\text{within}} = 0.245$; $k = 25$), was 0.075. This value suggests a small clustering effect at the class level, indicating that posttest self-efficacy scores were not strongly influenced by class membership.

In conclusion, the analysis found no significant difference in posttest self-efficacy between the ChatGPT and teacher-led groups after controlling for pretest scores, and the pretest covariate was not a significant predictor of posttest outcomes. To further verify the validity of these findings, the intraclass correlation coefficient (ICC) for posttest self-efficacy was examined and found to be small, indicating that class membership did not substantially influence the results. This minimal clustering effect supports the appropriateness and reliability of the ANCOVA analysis, even when working with two intact classes. Overall, these findings confirm that neither facilitation mode nor class-level factors had a major impact on self-efficacy development in this context.

Autonomy. Considering the autonomy pretest, homogeneity of variances

between autonomy levels of the two groups was assessed using Levene's Test, which was not significant ($p = .920$), indicating that the assumption of equal variances was satisfied. The intraclass correlation coefficient (ICC) was subsequently calculated using the one-way ANOVA method, with class as the grouping factor. Based on the ANOVA results ($MS_{\text{between}} = 0.034$; $MS_{\text{within}} = 0.086$; $k = 25$), the ICC was -0.025 , suggesting that there was no meaningful clustering effect at the class level for this outcome.

Although the intraclass correlation coefficient (ICC) confirmed that class-level clustering was not problematic for the analysis, the assumption of homogeneity of regression slopes was violated. The homogeneity of regression slopes assumption was tested by including an interaction term between groups and the covariate (Pretest). The interaction was statistically significant, $F(2, 47) = 4.138$, $p = .022$, indicating that the relationship between the covariate and the dependent variable differed significantly across groups. As a result, the assumption was violated, and ANCOVA was not deemed appropriate for this analysis. In other words, while the overall posttest scores did not show excessive clustering within groups, the predictive value of pretest scores on posttest outcomes was fundamentally altered in at least one group, reflecting a deep change in learning dynamics and the role of baseline knowledge. This pattern is particularly noteworthy because nearly all preliminary checks, such as normality, homogeneity of variance, and intraclass correlation, supported the use of ANCOVA. Nevertheless, the absence of a significant covariate effect contradicts statistical expectations. This result cannot be fully explained by quantitative data alone. It highlights the complexity of educational interventions in which deep changes in learner cognition, agency, or group dynamics may occur. These changes are not always captured by standard pretest measures. Therefore, to interpret these unexpected findings and reveal the underlying mechanisms, it is essential to integrate qualitative data. Qualitative analysis provides crucial insight into the ways participants' experiences and perspectives may have fundamentally changed over the course of the intervention.

Since there were only two groups and the assumption of homogeneity of regression slopes was not met, ANCOVA could not be appropriately applied. Therefore, we compared posttest scores between the two groups using an independent samples t-test. This alternative approach allows for a valid assessment of group differences in the outcome variable without relying on the covariate. All necessary assumptions for the t-test, including normality and homogeneity of variance, were checked before conducting the analysis.

Table 5. Autonomy posttest descriptive statistics.

	ClassID	N	Mean	Std. deviation	Std. error mean
Posttest	1	25	3.3280	.38033	.07607
	2	25	3.6520	.39475	.07895

Table 6. Autonomy posttest independent t-Test analysis

		Levene's test for equality of variances		t-test for equality of means					95% Confidence interval of the difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	Lower	Upper
Posttest	Equal variances assumed	.114	.737	-2.955	48	.005	-.32400	.10963	-.54443	-.10357
	Equal variances not assumed	.114	.737	-2.955	47.934	.005	-.32400	.10963	-.54444	-.10356

An independent samples t-test was conducted to compare posttest scores between the two groups. Levene's test indicated that the assumption of equal variances was met, $F = 0.114$, $p = .737$. The analysis revealed a statistically significant difference in posttest means, $t(48) = -2.955$, $p = .005$. The mean difference between groups was -0.324 , with a 95 percent confidence interval ranging from -0.544 to -0.104 . This result indicates that the group means differed significantly at the posttest stage. The independent samples t-test revealed a large effect size (Cohen's $d = 0.84$), indicating a substantial practical difference between the two groups. According to Cohen's widely cited benchmarks, an effect size of 0.8 or above is considered large, representing a strong separation between group means relative to within-group variation. Such a magnitude suggests that the intervention or condition under investigation yielded not only statistically significant but also meaningful real-world impacts on the measured outcome. In educational research, a large effect size at this level is relatively uncommon and points to the intervention's robust efficacy in shifting learner outcomes.

The intraclass correlation coefficient (ICC) for autonomy posttest, calculated using the one-way ANOVA method ($MS_{\text{between}} = 1.312$; $MS_{\text{within}} = 0.150$; $k = 25$), was 0.237. This value indicates a substantial clustering effect at the class level, suggesting that posttest autonomy scores were strongly influenced by class membership. The pronounced group effect observed in ANCOVA may thus be attributed in part to the emergence of class-level differences in autonomy following the intervention.

The quantitative results reveal a noteworthy shift in the pattern of learner autonomy after the intervention. Initially, the intraclass correlation coefficient (ICC) for autonomy was very low, indicating that autonomy was primarily an individual characteristic unrelated to class membership or context. However, after participating in project-based learning, the ICC increased substantially, reflecting a pronounced clustering effect at the class level. This suggests that autonomy, as measured at posttest, became strongly influenced by group dynamics and environmental factors rather than individual baseline levels.

Because these interpretations are derived from quantitative data alone, they should be treated as working assumptions. The observed clustering effect and the lack of covariate significance indicate that further qualitative investigation

is needed. The next stage of the research will incorporate focus-group interviews to clarify the mechanisms by which group and environmental factors shape the development of self-efficacy and autonomy in project-based self-learning settings. This mixed-methods approach will help ensure a comprehensive understanding of the complex dynamics at play.

Qualitative findings

To further understand the causes of similarities in efficacy and differences in autonomy levels, and especially to explore why pretest scores did not predict posttest outcomes, the research group conducted a thematic analysis of participant experiences. Analysis of both groups produced nine main themes. Seven themes were found in both groups. These are metacognitive skills, beliefs about the facilitator's role, motivation and desire, freedom and opportunity, metacognitive knowledge, receptive self-efficacy, and productive self-efficacy (see Appendix 1 and 2). Two other themes, peer negotiation, agency, and safety to fail, also appeared in both groups but showed differences in how they were expressed. The teacher-led group had one unique theme called delegated agency. This theme describes students relying on the teacher to make decisions. In total, there were twenty-seven subthemes. The coding combined theory-based categories with new ideas that appeared from the data.

Metacognitive skills. In the ChatGPT-facilitated group, students described a clear transition from passive to active planning. One participant noted, "We decided to throw out everything... creating our own set of rules" [C3]. This subtheme of the self-directed structure was confirmed by others, who reported "spending an entire meeting... not on the project's topic, but on making our own rules" [C1, C4, C7, C9]. Collaborative question design also emerged as the group agreed that "no one could approach the AI without the group first discussing and agreeing... questions were precious resources" [C3, C8]. For some, mastering this process became a key goal: "A really well-designed question... was like a master key" [C5, C10]. Reflective group decision-making was frequent, with students describing debates as "creative... fighting for ideas we were genuinely passionate about" [C5, C6, C9]. Most participants valued these changes, though a few, such as C2 and C7, participated less in group discussion and mainly followed. Several students later reflected that this was the first time they realized metacognitive skills could be self-determined, not just routines to satisfy the teacher. Some (C6, C8) described the process as waking up to how much classroom habits had previously limited their involvement and creativity. "I used to think autonomy just meant working alone and finishing my tasks, but now I see it is about building the process together and really taking charge as a group" [C8]. This reflective awareness marked a turning point, as students recognized their capacity to co-construct the process rather than simply comply with expectations.

In the teacher-led group, reliance on teacher-set routines was the norm. One student reported, "The first day always involved the teacher handing us a

detailed plan... my job was just to walk it" [T1]. The group's first actions were usually guided by the teacher rather than their ideas. Subthemes of limited independent planning and reflecting were noted to meet the teacher's standard: "We checked our product with the teacher's example... that is how we knew we were right" [T2, T5]. Debates or creative group reflection were rare. Most members accepted this routine, and there was little evidence of students suggesting changes or challenging the process. Several, however, reflected that the project made them more aware of how they had been conditioned to expect direction rather than agency (T6, T9).

Overall, the ChatGPT group's subthemes appeared much stronger, such as active planning, collaborative strategy, and reflective debate. In contrast, the teacher-led group's metacognitive activity was limited to following directions and seeking teacher approval, with minimal initiative or shared decision-making. This divergence suggests that the ChatGPT-facilitated environment fostered deeper metacognitive engagement and transformed students' conceptualization of autonomy.

Beliefs about the facilitator's role. Beliefs about the facilitator's role showed a clear contrast between groups. In the ChatGPT-facilitated group, students gradually moved from expecting teacher intervention to trusting group discussion and AI input. One participant described the shift: "We were looking around for the teacher, but the teacher was just observing... I said... 'the whole point is that there are no instructions'" [C4]. This led many to accept more responsibility, with several referring to the process as "taking charge ourselves" or "learning to decide as a group" [C1, C2, C4, C5, C8]. At the same time, a few still wished for traditional support: "Sometimes, I wanted a real person to tell me if my idea was good... I missed having someone give emotional support" [C2, C6]. The group valued the non-judgmental aspect of the AI, describing ChatGPT as "a partner" that "let us take risks... and not feel embarrassed" [C1, C7, C8, C10]. Most participants agreed with this new dynamic, though some, such as C5 and C9, were less vocal in decisions. For many, this shift in beliefs was transformative. Several (C4, C8) described stepping outside the panopticon, realizing for the first time how constant supervision and habitual authority had shaped their sense of agency. "I used to believe that only the teacher could tell me if I was doing the right thing, but now I see that our group can build the path together" [C4]. Some students explicitly stated that they had never experienced true group autonomy before this project-based learning model, and this realization completely changed how they perceived their own agency. This realization prompted deeper questioning of classroom norms and encouraged students to view autonomy as genuine group responsibility, not just individual independence.

In contrast, the teacher-led group consistently looked to the teacher for guidance and final decisions. Many described the teacher as the "final judge" or "the one who settled everything" [T2, T3, T4, T7, T8]. "The teacher would come over, listen for thirty seconds, and give us the answer... We got very good at giving up on disagreements, not at resolving them" [T2, T6]. Decision-making

was typically given over to the teacher, as shown by comments like “Let’s just present both ideas to the teacher and let her decide” [T2, T10]. Students valued encouragement from the teacher, but it reinforced the pattern of authority. The teacher was described as “very kind” but always led the group to safer choices [T3, T4]. There were no reports of students challenging this arrangement. This pattern demonstrates that authentic agency and collective responsibility were only realized when external control was minimized.

Motivation and desire. Motivation and desire were expressed very differently between the two groups. In the ChatGPT-facilitated group, motivation often shifted from grades to group creativity and enjoyment. One participant described, “When we made the whole class laugh with ChatGPT jokes... it was the first time I wanted to try harder not for a grade, but because it felt good to be creative” [C3, C9]. Several members used phrases like “fun,” “energy,” and “trying new things” to explain why they became more engaged [C2, C5, C8]. Others described the process as “learning a superpower” when they mastered prompt design [C5, C7]. Group motivation grew as ideas became more creative and collaborative. No one reported working mainly for grades, and most agreed that “contributing ideas” and “making each other laugh” became important goals [C1, C4, C10]. Some described this new orientation toward motivation as a transformative experience, where group creativity and intrinsic enjoyment replaced a sole focus on grades. Some students (C3, C8, C10) shared that for the first time, they felt real ownership of their learning, challenging their previous assumptions about what success and engagement in class should mean. “Before, I thought being motivated was just about not losing points, but now I see real motivation comes from creating something new with others” [C10].

In contrast, the teacher-led group was mainly motivated by evaluation and teacher praise. One student stated, “My main concern is always the final grade. An experimental format is a gamble. A standard format, done well, is a guaranteed good score” [T5]. Others confirmed they felt happiest when praised by the teacher, describing it as “the real reward” [T3, T4, T6]. Some participants mentioned feeling “deflated” or “shut down” when their creative ideas were rejected in favor of safer, teacher-approved formats [T2, T8]. The group focused on “doing assigned tasks” and “avoiding mistakes” rather than taking creative risks. A few students (T6, T9, T10) noted that the experience made them question, for the first time, whether grades should be the only source of motivation. These differences indicate that the ChatGPT context shifted learners toward more intrinsic and collaborative forms of motivation, while the teacher-led context reinforced external validation.

Freedom and opportunity. Freedom and opportunity were experienced differently in the two groups. In the ChatGPT-facilitated group, students felt increasing freedom to experiment and try new approaches. One participant shared, “I could experiment with silly ideas without worrying about being judged. ChatGPT never got annoyed or impatient, so I felt free to try even things that might fail” [C2]. Many group members described having “room to

experiment” and “no fear of being wrong” [C1, C3, C4, C8, C10]. The group’s autonomy led them to create their own “Group Constitution,” setting rules for feedback and decision-making without teacher involvement [C4]. Several called this “the most valuable lesson” or “real autonomy” [C4, C5].

Some even noted it was the first time they became aware of the invisible boundaries that had structured their actions in previous classroom settings. In the teacher-led group, students reported limited freedom and considered risk-taking unsafe. One participant said, “We always had to check with the teacher before doing anything different; it was safer that way” [T2, T4]. Others described feeling “anxious” about trying new formats, and most preferred to stick to what the teacher expected [T3, T5, T8, T9]. The teacher assigned specific roles, and students mainly did only their part. “Not much chance to do more,” as one put it [T3]. Group discussions about changing plans were rare, and most participants described “staying within the safe path” [T1, T6, T7, T10]. The findings show that an authentic opportunity for innovation and ownership was mainly realized in the ChatGPT group, reflecting a significant difference in classroom ecology.

In summary, students in the ChatGPT group described greater freedom and real opportunities to shape their work, with several using words like “experiment,” “no fear,” and “autonomy.” In contrast, the teacher-led group operated within fixed limits, with nearly all participants confirming that teacher control and the desire to avoid mistakes made them less likely to take risks or try new ideas.

Metacognitive knowledge. Metacognitive knowledge showed strong contrasts between the two groups. In the ChatGPT-facilitated group, students described learning new strategies and monitoring their progress more closely. One participant explained, “I copy and pasted the first paragraph into the AI and wrote, ‘Explain this to me like I am a high school student’... bit by bit, I dismantled that wall... I proved to myself that I was capable of understanding difficult things” [C2]. Others confirmed that using the AI for research and clarification helped them “feel prepared” and “make my voice feel solid and credible” in group discussions [C1, C3]. The group valued “evidence-based contribution.” Several members reported greater confidence in sharing or challenging ideas based on their investigation.

In the teacher-led group, metacognitive knowledge was mainly tied to external correction and comparison with teacher examples. One student said, “I was so worried about doing something wrong that I would finish my part and then just sit there... waiting for the teacher to have a free moment to come and check it” [T3]. Participants described “comparing my part to the example” and “changing it to match,” rarely reflecting on their learning process or trying new approaches [T2, T4]. Most admitted they focused on “finishing on time” and “waiting for approval” rather than understanding or improving how they learned.

Overall, students in the ChatGPT group described building new strategies, monitoring their progress, and confidently sharing research-based opinions.

The teacher-led group relied on external checks and corrections, with most participants confirming they focused on matching the teacher's model rather than developing their learning strategies.

Receptive self-efficacy. Students in the ChatGPT-facilitated group described a clear improvement in their ability to understand difficult material. One participant shared, "My ability to read complex texts completely changed... I used the method our group had developed" [C2]. Several others described feeling "capable of understanding difficult things" and less intimidated by academic readings [C1, C3]. Many noted that working with group strategies and the AI helped them become more "resourceful" and independent.

In the teacher-led group, receptive self-efficacy mainly depended on teacher support. One student said, "The rehearsal with the teacher was perfect... We felt very confident going into the final presentation" [T5]. Most participants explained that if they did not understand a reading, they waited for the teacher to explain it and rarely tried to figure it out themselves [T3, T4]. Words like "waiting" and "relief" were common.

In summary, ChatGPT-group built confidence through their strategies and peer support, while teacher-led students mainly relied on the teacher for understanding and reassurance. Most in the teacher-led group confirmed they did not attempt to solve comprehension problems independently.

Productive self-efficacy. Productive self-efficacy, or confidence in performing and expressing ideas, developed differently across the two groups. In the ChatGPT-facilitated group, students gained confidence by practicing with unpredictable scenarios. One explained, "The judges' questions felt exactly like the hard questions we had been practicing with the AI for weeks. It felt familiar" [C5, C7, C10]. Others noted that group support helped them handle unexpected problems during presentations. "We just kept going... it was not a mistake; it was a duet. We were working together, live, without a script" [C4, C6]. Most students described feeling "more adaptable" and less anxious about making mistakes in front of others [C1, C3, C8, C9]. Several students described this as a transformative change. The ability to adapt and improvise in real time reshaped their concept of confidence and made them realize that productive self-efficacy is more about flexibility and authentic performance than mere repetition. "I never thought before that real confidence comes from handling surprises, not just repeating what I practiced. Now I am much less afraid of making mistakes in public" [C7].

In the teacher-led group, productive self-efficacy was mostly linked to rehearsed scripts and teacher instructions. "I practiced my part over and over the way she told me... I just repeated what I learned, which made me feel safe" [T1, T2, T3]. However, when faced with unscripted questions, students often froze. "I completely froze... all that practice and I couldn't handle one unexpected moment" [T5, T8]. Most students said their confidence was high only when everything went according to plan [T4, T6, T7, T9, T10]. These findings

highlight that the ChatGPT group developed greater adaptability and resilience, while the teacher-led group depended on predictability.



Peer negotiation and agency. Peer negotiation and agency were much more visible in the ChatGPT-facilitated group. Students often resolved conflicts themselves, discussing problems openly as a team. One participant recalled, “It got quiet and tense... Finally, after one particularly bad meeting, a few of us stayed behind and just said, ‘We can’t continue like this’... We asked ChatGPT for advice... The next day, one of us took a deep breath and started the conversation” [C4, C8]. Others described creating a group constitution to formalize how decisions and feedback would be managed [C3, C6, C9, C10]. Many students highlighted building “trust” and “support” within the group and valued the chance to “help each other when we were stuck” [C1, C2, C5, C7]. No one reported relying on outside authority for group decisions. For many, the opportunity to negotiate and challenge group norms for the first time was eye-opening. “I used to just go along with whatever was decided, but this time I learned to really speak up and shape what the group does. It changed my sense of what agency really means” [C9]. This shift fostered critical awareness of how agency could be shared and enacted collectively, rather than simply granted or withheld by authority figures.

In the teacher-led group, peer negotiation was rare and often replaced by teacher intervention. “Instead of trying to blend them or debate the merits of each, they both just ran to the teacher separately and asked, ‘Which one is better?’... The teacher... just pointed at one and said, ‘Use this one’” [T5, T9]. Students said it was “hard to change group dynamics without her help” and most “just did our tasks quietly, not much teamwork except when the teacher told us to discuss” [T2, T3, T6, T7, T8]. Few reported trying to solve their disagreements or conflicts [T1, T4, T10]. In summary, actual group negotiation and agency only emerged under conditions where authority was decentralized.

Safety to fail. The sense of safety to fail was experienced very differently between the two groups. In the ChatGPT-facilitated group, students described an environment where mistakes were accepted and treated as part of the learning process. One participant shared, “I could ask the same thing five times, try different versions, and no one would laugh” [C2]. Others said “mistakes became lessons, not something to be punished” [C3]. Practicing with ChatGPT allowed students to “rehearse in private” and “try again in the group” without fear of embarrassment [C5]. Many agreed that this environment encouraged risk-taking and made it easier to “catch a teammate” during group work [C4].

In contrast, the teacher-led group described a much lower sense of psychological safety. “My biggest fear has always been looking stupid. Making a mistake in front of the teacher or even my group members felt like a public failure. So, I developed a habit of silence” [T3]. Most students reported “avoiding risks,” being “scared to write” without correction, and preferring not to speak unless they were certain. Corrections were seen as necessary, but also made students more dependent on the teacher’s judgment [T2, T4].

Overall, the ChatGPT group described an atmosphere where failure was safe and useful, while the teacher-led group associated mistakes with embarrassment or risk, leading to more caution and silence in class.

Delegated agency: a unique theme of the teacher-led group. Delegated agency was a unique theme found only in the teacher-led group. Students often handed over decision-making to the teacher instead of negotiating within the group. One participant described, “The argument didn’t end with us coming to a decision. It ended when Mai said, ‘Let’s just present both ideas to the teacher and let her decide’” [T2]. Many called this “a relief” or “easier,” as it removed the pressure to resolve conflicts themselves [T1, T3]. Several students admitted that in most projects, “the first day always involved the teacher handing us a detailed plan... my immediate thought was always one of relief: ‘Okay, as long as I follow these steps exactly, I can’t go wrong’” [T1]. Few reported challenging this habit, and most were comfortable letting the teacher make final decisions.

This theme did not appear in the ChatGPT-facilitated group, where students described taking ownership of group choices and rarely seeking outside intervention. In the teacher-led group, however, the delegated agency was the norm and strongly shaped the learning environment. Most participants confirmed that they preferred to let the teacher decide rather than work through disagreements themselves.

Across all themes, the project enabled many students to engage in critical and transformative reflection, leading to new awareness of classroom power, surveillance, and genuine autonomy. Many students explicitly stated that they previously misunderstood autonomy as mere independence or compliance, and only through this model did they come to a new, more profound understanding of what autonomy and agency mean. This shift helps explain why posttest measures diverged so sharply from pretest expectations and why the pretest covariate could not predict posttest outcomes. The learning experience itself catalyzed a reconsideration of what it means to have and enact agency in a learning environment.

Discussion

RQ1. ChatGPT-facilitated project-based learning fundamentally transforms learner autonomy in EFL contexts.

This study demonstrates that ChatGPT-facilitated project-based learning produces significantly greater gains in learner autonomy than traditional teacher-led PBL. Importantly, the pretest autonomy score did not predict posttest autonomy, which is an atypical finding in educational research. This phenomenon is not a result of methodological error. Rather, it is a recognized feature of group-based interventions such as project-based learning. Collaborative learning environments foster close peer interaction, the negotiation of group norms, shared strategies for self-regulation, and mutual support. These processes can lead to the emergence of distinct classroom cultures of autonomy,

where group-level influences outweigh individual differences present before the intervention. Additional factors, such as teacher facilitation style, physical and technological resources, and the social history of each class, may also contribute to these unique ecological contexts.

The increase in the intraclass correlation coefficient (ICC) from pretest to posttest indicates that, following the intervention, students within the same group became more similar to each other in their outcome scores. This pattern is consistent with the development of stronger group cohesion or shared practices during collaborative learning. Such clustering effects are often observed when members engage in sustained interaction and collective problem-solving, resulting in a group-specific learning culture or sense of agency that is not present at baseline.

Based on the qualitative data, in the ChatGPT group, students took initiative in planning, designing questions, and negotiating group roles. These behaviors embody the principles of autonomy and distributed agency that have been identified in prior scholarship (Panadero, 2017; Benson, 2011). The findings confirm that, when digital mediation is combined with explicit project structure and orientation, learners can benefit from both freedom and sufficient scaffolding, which prevents disorientation (Lee & Hannafin, 2016; Nguyen, 2020). Notably, this contrasts with previous research suggesting that lower-proficiency EFL learners struggle with reduced guidance (Kirschner et al., 2006).

The underlying mechanism behind these autonomy gains is both procedural and conceptual. Previous studies have highlighted the limiting role of teacher authority in Asian EFL classrooms (Littlewood, 1999; Chan, 2001); however, the structured use of AI in this study transformed such cultural expectations and enabled authentic peer-driven learning. Qualitative data reveal that students did not only act more autonomously, but they also redefined what autonomy meant. Many participants described a transition from passive compliance to active, collective agency. Several students explicitly reported that, before the intervention, they believed autonomy meant completing assigned work individually. Through total group self-regulation, they discovered that true autonomy involves co-constructing rules, negotiating disagreements, and taking shared ownership of outcomes. This process exemplifies a shift in learner perspective, in which students revise core assumptions through direct and reflective experience (Mezirow, 1997; Illeris, 2014; Hoggan, 2016).

This profound shift in understanding was also observable in the quantitative results. The sharp increase in the intraclass correlation coefficient (ICC) for autonomy after the intervention indicates that group-level agency became a real and measurable phenomenon (Smith & Erdoğan, 2008; Benson, 2011). These findings underscore the power of AI-facilitated, project-based learning to catalyze individual and collective transformation in student agency, even within highly teacher-centered educational cultures.

RQ2. Both ChatGPT and teacher-led facilitation foster self-efficacy, but through distinct mechanisms.

Both groups in this study experienced gains in self-efficacy, although the mechanisms underlying these gains were different. The ChatGPT group built self-efficacy through repeated, low-risk experimentation and immediate feedback from the AI, consistent with Bandura's (1997) emphasis on mastery experience and the value of supported practice. Recent research further supports the role of AI in lowering anxiety and encouraging productive risk-taking (Gayed, 2025; Xu et al., 2024; Yildiz, 2024). By contrast, the teacher-led group relied more on structured encouragement and direct guidance, reflecting the socio-emotional support described by Mercer and Dörnyei (2020).

Interestingly, the quantitative analysis revealed no significant difference in overall self-efficacy gains between the two groups. This result differs from some studies that report stronger effects for teacher support among lower-proficiency learners (Hmelo-Silver et al., 2007). In this study, scaffolding of the project in the human teacher-led class, constant support from ChatGPT in the Chat-GPT-as-the-facilitator class, and the use of explicit rubrics, the orientation phase likely reduced uncertainty and supported both groups equally. These results suggest that both digital and human facilitation can foster self-efficacy, provided that project-based approaches and immediate support are present.

RQ3. Project-based self-regulation with AI catalyzes a fundamental redefinition of autonomy and agency.

Qualitative analysis shows that total self-regulation in the ChatGPT-facilitated group changed students' observable behaviors and produced a deep transformation in their beliefs about autonomy. Many students explicitly stated that, prior to this experience, they had considered autonomy as independent work by simply following instructions. When given the opportunity to co-create group norms, resolve conflicts collaboratively, and experience freedom from teacher surveillance, students begin to question and then revise their understanding of what it means to be an autonomous learner. Several participants described this as "realizing the classroom had invisible boundaries" and "discovering we could actually decide things ourselves." These outcomes are consistent with transformative learning theory's core principles, which emphasize that perspective transformation is central to educational change that prevents the pretest from predicting the posttest results (Mezirow, 1997; Illeris, 2014; Hoggan, 2016).

In particular, this redefinition of autonomy and agency helps to explain why pretest autonomy scores did not predict posttest outcomes, and why class-level clustering of autonomy (as indicated by ICC) emerged only after the intervention. The findings confirm that authentic, collective agency and a new ecology of learning become possible only when students experience and internalize self-regulation in practice, rather than viewing it as an abstract ideal (Benson, 2011; Lee & Hannafin, 2016; Littlewood, 1999; Chan, 2001).

Conclusion

This study provides robust evidence that AI-facilitated self-regulated project-based learning, implemented through ChatGPT, can fundamentally transform learner autonomy and self-efficacy in EFL contexts traditionally dominated by teacher-centered norms. Students in the ChatGPT group achieved significantly greater autonomy gains than those in teacher-led groups, and more importantly, developed a qualitatively different, more reflective, and collaborative understanding of autonomy itself. This transformation is reflected in both the marked posttest increase in the intraclass correlation coefficient (ICC) for autonomy and the qualitative evidence of students moving from compliance and dependence to authentic self- and co-regulation (Mezirow, 1997; Illeris, 2014; Benson, 2011).

Both ChatGPT- and teacher-led facilitation produced significant improvements in learner self-efficacy, though via distinct pathways. The AI-supported group gained confidence through iterative experimentation and peer feedback, while the teacher-led group benefited from direct guidance and structured reassurance (Bandura, 1997; Mercer & Dörnyei, 2020). The lack of significant between-group differences in self-efficacy gains, along with the limited predictive value of pretest scores, suggests that clear project design and structured scaffolding are essential for supporting learner growth, regardless of facilitation mode.

Crucially, this study shows that project-based self-regulation with AI not only alters observable behaviors but also leads to a fundamental redefinition of self-regulation and autonomy in learning. Qualitative data illustrate how participants recognized, questioned, and reconstructed their views on classroom authority, collaboration, and self-directed learning. This process of perspective transformation aligns with core tenets of transformative learning theory and explains why initial differences in autonomy did not predict posttest outcomes (Mezirow, 1997; Hoggan, 2016).

Some methodological limitations must be acknowledged. The research was limited to a single language center with a modest sample size and randomization at the class, not individual, level. Although group comparability and clustering were addressed, factors such as novelty effects, digital literacy, and prior AI experience may have affected the results. The measures used, while internally consistent, require further validation across broader contexts with more designed measurement validation tools.

Despite these limitations, the findings provide initial and exploratory evidence that AI can catalyze both individual and collective transformation in autonomy, even within educational settings long shaped by teacher authority. The study underscores that the effectiveness of educational technology depends not only on the tool itself but also on its integration with explicit project design, reflective practice, and the willingness to disrupt traditional classroom routines. Future research should expand this work by involving larger and more diverse samples, using validated digital literacy measures, and adopting longitudinal designs to examine the durability of these effects. Additional studies should also investigate how project complexity, facilitation

style, and classroom culture influence the impact of AI on learner agency and self-efficacy.

In summary, this exploratory study demonstrates that AI-facilitated project-based learning can both enhance and fundamentally redefine learner autonomy in EFL education. These findings highlight the need for educational policy and practice to embrace and support reconceptualization of learner roles, technology-mediated learning environments, as the landscape of language education evolves.

References

- Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(2), 39–43. <https://doi.org/10.1080/00098650903505415>
- Benson, P. (2011). *Teaching and researching autonomy in language learning* (2nd ed.). Routledge.
- Biesta, G. (2020). Risking ourselves in education: Qualification, socialization, and subjectification revisited. *Educational Theory*, 70(1), 89–104. <https://doi.org/10.1111/edth.12411>
- Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3–4), 369–398.
- Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, 8(1), 102418. <https://doi.org/10.29140/ajal.v8n1.102418>
- Chan, V. (2001). Readiness for learner autonomy: What do our learners tell us? *Teaching in Higher Education*, 6(4), 505–518.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Dörnyei, Z., & Murphey, T. (2003). *Group dynamics in the language classroom* (Vol. 10). Cambridge University Press.
- Duc, C. N., Thi, P. N., Hoang, T. N., & Thanh, T. N. T. (2022). A literature review of the project-based teaching method in the education of Vietnam. *International Journal of Educational Methodology*, 8(3), 567–584. <https://doi.org/10.12973/ijem.8.3.567>
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383. <https://doi.org/10.2307/2666999>
- Gayed, J. M. (2025). Educators' perspective on artificial intelligence: Equity, preparedness, and development. *Cogent Education*, 12(1), 2447169. <https://doi.org/10.1080/2331186X.2024.2447169>

- Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In G. Joughin (Ed.), *Tertiary assessment & higher education student outcomes: Policy, practice & research* (pp. 259–275). University of Wollongong.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107. <https://doi.org/10.1080/00461520701263368>
- Hoggan, C. D. (2016). Transformative learning as a metatheory: Definition, criteria, and typology. *Adult Education Quarterly*, 66(1), 57–75. <https://doi.org/10.1177/0741713615611216>
- Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon Press.
- Holstein, K., McLaren, B. M., & Aleven, V. (2020). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *Educational Technology Research and Development*, 68, 1613–1634.
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Illeris, K. (2014). Transformative learning and identity. *Journal of Transformative Education*, 12(2), 148–163. <https://doi.org/10.1177/1541344614548423>
- Immordino-Yang, M. H., Darling-Hammond, L., & Krone, C. R. (2018). *The brain basis for integrated social, emotional, and academic development*. The Aspen Institute. <https://www.aspeninstitute.org/publications/the-brain-basis-for-integrated-social-emotional-and-academic-development/>
- Ivankova, N. V., Creswell, J. W., & Stick, S. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20.
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. (2023). Promoting socially shared regulation of learning and emotional awareness in collaborative learning. *British Journal of Educational Psychology*, 93(1), 257–273. <https://doi.org/10.1016/j.chb.2015.03.082>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Kutuk, G., Putwain, D. W., Kaye, L. K., & Garrett, B. (2023). The development and preliminary validation of a new measure of self-efficacy: Questionnaire of Self-Efficacy in Learning a Foreign Language (QSL). *ITL – International Journal of Applied Linguistics*, 174(4), 506–535. <https://doi.org/10.1075/itl.21031.kut>
- Lee, E., & Hannafin, M. J. (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational Technology Research and Development*, 64(4), 707–734. <https://doi.org/10.1007/s11423-015-9422-5>

- Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 14–29. <https://doi.org/10.2167/illt040.0>
- Littlewood, W. (1999). Defining and developing autonomy in East Asian contexts. In S. Cotterall & D. Crabbe (Eds.), *Learner autonomy in language learning: Defining the field and effecting change* (pp. 77–88). Peter Lang.
- McIntyre, N. A. (2016). Teach at first sight: Expert teacher gazes across two cultural settings (Doctoral dissertation, University of York). <https://etheses.whiterose.ac.uk/id/eprint/15408/>
- Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press.
- Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education*, 1997(74), 5–12. <https://doi.org/10.1002/ace.7401>
- Nghia, V. N. D., & Nguyen, Q. N. (2021). Task-based language teaching to enhance learner communicative competence in Vietnam: A matter of opinion. *International Journal of Linguistics, Literature and Translation*, 4(8), 151–159. <https://doi.org/10.32996/ijllt.2021.4.8.22>
- Nguyen, S. V., & Habók, A. (2021). Designing and validating the learner autonomy perception questionnaire. *Heliyon*, 7(4), e06831. <https://doi.org/10.1016/j.heliyon.2021.e06831>
- Nguyen, Q. N. (2022). Teachers' scaffolding strategies in internet-based ELT classes. *TESL-EJ*, 26(1), n1. <https://doi.org/10.55593/ej.26101a1>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8(422). <https://doi.org/10.3389/fpsyg.2017.00422>
- Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 15–31). Academic Press. <https://doi.org/10.1016/B978-012750053-9/50003-6>
- Schwartz, M. (2024). *Ecological perspectives in early language education: Parent, teacher, peer, and child agency in interaction*. Taylor & Francis.
- Smith, R., & Erdoğan, S. (2008). Teacher-learner autonomy: Programme goals and student-teacher constructs. In T. Lamb & H. Reinders (Eds.), *Learner and teacher autonomy* (pp. 83–102). John Benjamins Publishing Company.
- Thomas, J. W. (2000). A review of research on project-based learning. *The Autodesk Foundation*. http://www.bobpearlman.org/BestPractices/PBL_Research.pdf
- Thomas, M. S. C., Ansari, D., & Knowland, V. C. P. (2019). Annual research review: Educational neuroscience: Progress and prospects. *Journal of Child Psychology and Psychiatry*, 60(4), 477–492. <https://doi.org/10.1111/jcpp.12973>
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.

- Wu, Z., Abdul Halim, H., & Mohd Saad, M. R. (2025). Impact of AI-supported multimodal learning environments on EFL students' critical thinking and self-regulation: A quasi-experimental study in Shanxi Province, China. In *Proceedings of the International Conference on Modern Research in Education, Teaching and Learning* (Vol. 4, No. 1, pp. 23–33).
- Xu, S., Chen, P., & Zhang, G. (2024). Exploring the impact of the use of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment. *Heliyon*, 10(21), e23949. <https://doi.org/10.1016/j.heliyon.2024.e39845>
- Yildiz, C. (2024). ChatGPT integration in EFL education: A path to enhanced speaking self-efficacy. *Novitas-ROYAL (Research on Youth and Language)*, 18(2), 167–182. <https://novitasroyal.org/index.php/novitas/article/view/2213>
- Zhang, Q., Nie, H., Fan, J., & Liu, H. (2025). Exploring the dynamics of artificial intelligence literacy on English as a foreign language learners' willingness to communicate: The critical mediating roles of artificial intelligence learning self-efficacy and classroom anxiety. *Behavioral Sciences*, 15(4), 523. <https://doi.org/10.3390/bs15040523>
- Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using Gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, 21(2), 2175. <https://doi.org/10.29140/jaltcall.v21n2.2175>
- Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. *Language Learning & Technology*, 29(1), 1–29. <https://hdl.handle.net/10125/73606>