

Agentic AI as Process Scaffolding in Undergraduate ESL Instruction: Effects on Proficiency, Engagement, and Self-Regulated Language Learning

Roland A. Niez

Biliran Province State University, Naval, Biliran, Philippines

Joedel O. Peñaranda*

Biliran Province State University, Naval, Biliran, Philippines

Maribel N. Zipagan

Biliran Province State University, Naval, Biliran, Philippines

John Cavin M. Sabonsolin

Biliran Province State University, Naval, Biliran, Philippines

Cristie Ann L. Jaca

Cebu Technological University, Cebu City, Philippines

Noel P. Tancinco

Biliran Province State University, Naval, Biliran, Philippines

Jelyn M. Magno

Biliran Province State University, Naval, Biliran, Philippines

Mary Joy B. Gelizon

Biliran Province State University, Naval, Biliran, Philippines

Jed Priam Cerera

Biliran Province State University, Naval, Biliran, Philippines

Abstract—Recent AI-assisted language learning research has focused mainly on reactive chatbots that respond to prompts but offer limited support for planning, monitoring, and follow-up. This study examined whether a bounded agentic AI environment designed as process scaffolding would be associated with stronger ESL outcomes than a reactive AI interface. Using an explanatory sequential mixed-methods, quasi-experimental design, 128 undergraduates from four intact academic communication sections at two private universities in the Philippines participated in a 12-week intervention. Students in the experimental condition used an Agentic English Coach that supported weekly goal setting, task decomposition, reminders, reflection prompts, adaptive resources, and teacher escalation; the comparison group used a reactive chat interface to complete the same course tasks. ANCOVA results controlling for pretest scores showed that the agentic AI group outperformed the reactive AI group in posttest English proficiency, learner engagement, self-regulated language learning, and agentic engagement. Within the experimental group, plan completion, reflection completion, revision cycles, and speaking practice turns positively predicted posttest proficiency. Interview data indicated that agentic AI reduced task uncertainty, supported sustained practice, made progress more visible, and shifted teachers toward orchestration and oversight of feedback rather than constant procedural prompting. The findings suggest that AI is most pedagogically valuable in ESL instruction when it structures learning processes over time rather than only generating on-demand responses.

Index Terms—agentic artificial intelligence, ESL instruction, learner engagement, self-regulated language learning, language pedagogy

* Corresponding Author. Email: jpenaranda@bipsu.edu.ph

I. INTRODUCTION

Generative artificial intelligence is moving quickly from an emerging technology to a routine part of language education. In English language classrooms, this shift is especially visible because drafting, rehearsal, revision, and feedback can all be supported through conversational interfaces. Recent research shows that AI tools can enhance English learning achievement, participation, and motivation when their use is tied to clear pedagogical tasks and teacher guidance (Kasneci et al., 2023; Wei, 2023; Xu & Wang, 2024; Wu, 2024). In ESL and EFL settings, AI-mediated practice has also been associated with gains in speaking performance, willingness to communicate, and academic engagement (Fathi et al., 2024; Wang & Xue, 2024; Zhang et al., 2024).

At the same time, most AI applications in language learning remain reactive. Learners can ask a question, submit a prompt, or request a revision, and the system responds to that request immediately. Although this form of support can be useful, it often leaves the burden of planning, pacing, monitoring, and follow-up almost entirely with the learner. That is a significant pedagogical limitation because successful language learning rarely depends on isolated answers alone. It is more often sustained by sequences of forethought, practice, feedback, strategic adjustment, and reflection. For many learners, especially in undergraduate ESL classrooms where proficiency, confidence, and study habits vary considerably, the difficulty lies not only in producing language but also in organizing the process by which improvement occurs.

The emerging concept of agentic AI offers a potentially important shift. Agentic systems are distinguished not merely by their ability to generate language, but by their ability to pursue bounded goals through planning, memory, monitoring, and coordinated action across multiple steps (Plaat et al., 2025). Rather than functioning only as answer engines, they can act as learning partners that help learners set subgoals, track progress, recommend next actions, and reroute activity when difficulties arise. In educational terms, this matters because it connects AI not just to content support, but to the architecture of learning behavior itself. Molenaar (2022) argues that effective human-AI regulation depends on productive coordination between learner control and intelligent scaffolding. Likewise, Liu et al. (2025) show that AI-mediated English learning can foster agentic practices when learners use digital tools to make purposeful decisions about their own progress.

This perspective is especially relevant to learner engagement and self-regulated language learning. Engagement has traditionally been conceptualized as behavioral, emotional, and cognitive involvement in learning (Fredricks et al., 2004), yet later work highlights a fourth dimension, agentic engagement, in which learners actively shape instruction rather than simply respond to it (Reeve & Tseng, 2011). Self-regulated learning, meanwhile, emphasizes the learner's capacity to set goals, select strategies, monitor performance, and reflect on outcomes (Zimmerman, 2002; Panadero, 2017). These constructs are deeply interconnected in ESL learning, where progress depends on repeated cycles of intention, strategic action, and adaptation. If agentic AI can make those cycles more visible and manageable, then its effects may be felt not only in final performance but also in the quality of learners' ongoing participation.

Recent research offers promising but still incomplete evidence. ChatGPT and related tools have been linked to self-learning routines among EFL learners, yet studies also note variability in the depth of use, the quality of strategies, and critical evaluation (Tram et al., 2024; Yang & Li, 2024). Other work suggests that AI can support self-regulated learning when systems are designed around autonomy, competence, and reflective prompting (Chiu, 2024; Xia et al., 2023). A systematic review by Guan et al. (2025) concludes that educational chatbots can scaffold planning, monitoring, and feedback, but that many implementations still focus on short-term interactions rather than long-horizon orchestration. In other words, the strongest theoretical promises of AI for language learning may depend less on the quality of one-time responses and more on whether the system can sustain purposeful learning trajectories over time.

The present study addresses this gap by comparing a bounded agentic AI environment with a reactive AI interface in undergraduate ESL instruction. The intervention was situated in a Philippine higher education context, where English functions both as a medium of academic work and as a resource for social mobility. Rather than asking whether AI is generally useful, the study examines a narrower pedagogical question: what changes when AI is designed to scaffold learning processes across time rather than answer prompts on demand? By analyzing English proficiency, learner engagement, self-regulated language learning, system-use patterns, and post-intervention perceptions, the study aims to clarify how agentic AI may reshape routines of practice, feedback, and teacher orchestration in contemporary ESL pedagogy.

II. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

A. From Reactive Chatbots to Agentic AI in Language Education

The distinction between reactive and agentic AI is critical for interpreting current developments in language education. Reactive systems respond to prompts users provide, but they typically do not maintain long-term instructional intent unless learners explicitly reconstruct it in each interaction. Agentic AI, by contrast, is oriented toward goal pursuit across time. Plaat et al. (2025) describe agentic large language model systems as architectures capable of planning, tool use, memory retention, and multi-step action under explicit constraints. From a pedagogical standpoint, these properties matter because learning tasks are not single events. A student writing an argumentative paragraph, preparing an oral presentation, or building vocabulary for a disciplinary topic must manage a sequence of interconnected decisions over time. A tool that can remember these goals, propose next steps, and monitor partial completion may therefore function differently from a

tool that simply waits for the next prompt.

Language education research has already begun to reveal the limits of purely reactive support. Yang and Li (2024) note that many current L2 applications of ChatGPT are strongest in feedback and idea generation but weaker in ensuring strategic progression, task continuity, and responsible use. Tram et al. (2024) likewise found that self-learning with ChatGPT could promote flexibility and convenience, yet learners' gains depended heavily on whether they adopted disciplined routines rather than sporadic prompt experimentation. In a formal ESL curriculum, that distinction becomes even more significant because progress must be aligned with course outcomes, assessment rubrics, and teacher expectations.

Agentic AI is also relevant to how pedagogy is distributed between human and machine. Molenaar (2022) proposes hybrid human-AI regulation, in which intelligent systems can support phases of self-regulation without displacing the teacher's interpretive and ethical role. That framing is useful for language classrooms because AI support is most beneficial when it extends rather than replaces pedagogical judgment. In an ESL setting, for example, a system may help learners schedule speaking rehearsals, flag repeated grammar errors, or prompt self-explanations after revision. However, teachers remain crucial for calibrating accuracy, contextualizing language use, and ensuring that learners do not outsource meaning-making itself.

The notion of agentic practices is especially salient in digitally mediated English learning. Liu et al. (2025) link AI-mediated informal digital learning of English to learners' critical digital literacies and agentic decision-making. Their findings suggest that AI becomes educationally powerful when learners use it to organize effort, negotiate strategies, and evaluate information rather than passively consume outputs. This insight implies that the pedagogical effectiveness of agentic AI depends not only on the system's technical sophistication, but also on whether classroom design positions learners as active managers of their own language development.

B. Agentic AI and Learner Engagement

Engagement remains one of the strongest predictors of language learning persistence and quality. Fredricks et al. (2004) conceptualize engagement as a multidimensional construct that includes behavioral participation, emotional involvement, and cognitive investment. Reeve and Tseng (2011) extend this framework by introducing agentic engagement, which refers to learners' constructive contribution to the flow of instruction through questions, preferences, and self-initiated input. In AI-supported language learning, this fourth dimension deserves special attention because the very presence of interactive technologies may either expand or suppress learner initiative depending on how the technology is designed.

Empirical findings in L2 education suggest that AI can increase engagement, but the mechanisms vary. Wang and Xue (2024) found that AI-driven chatbots can foster academic engagement in EFL classrooms when activities are structured and sustained across an intervention period. Fathi et al. (2024) reported gains in speaking skills and a willingness to communicate through AI-mediated interactions, suggesting that AI may reduce anxiety and increase opportunities for productive practice. Zhang et al. (2024) similarly showed that AI-supported speaking tools can help less confident learners move toward more active oral participation. These studies are encouraging, yet they mostly examine systems that respond during task performance rather than systems that proactively sustain engagement before, during, and after tasks.

Agentic AI may influence engagement through several pathways. First, it can lower initiation barriers by transforming broad assignments into manageable action steps. Second, it can make progress visible through reminders, dashboards, and checkpoints, which may encourage behavioral persistence. Third, it can invite learners to shape the learning sequence itself by selecting goals, revising plans, or requesting additional practice, thereby supporting agentic engagement. Reeve and Shin (2020) argue that teaching for agentic engagement involves creating structures that welcome learners' constructive input. An agentic AI system, when designed carefully, may operationalize that principle at scale by prompting learners to articulate preferences, make decisions, and respond to feedback loops more frequently than is feasible through teacher interaction alone.

Still, engagement gains should not be assumed automatically. Kasneci et al. (2023) warn that educational uses of large language models can lead to superficial compliance if learners rely on polished outputs without genuine cognitive investment. The issue is therefore not whether AI interacts with learners, but whether it does so in ways that deepen, rather than simulate, participation. This makes it important to study engagement not only as enjoyment or frequency of use, but as a pattern of visible effort, strategic input, and constructive initiative.

C. Agentic AI and Self-Regulated Language Learning

Self-regulated learning is central to language development because learners must manage distributed practice across classrooms, homework, revision, and informal exposure. Zimmerman (2002) organizes self-regulation into cyclical phases of forethought, performance, and self-reflection, while Panadero (2017) synthesizes multiple models that emphasize planning, monitoring, control, and evaluation. In ESL contexts, these processes are not optional extras. Learners who progress steadily tend to define language goals, allocate practice time, monitor breakdowns, seek feedback, and revise strategically. By contrast, learners who engage in fragmented or last-minute practice often struggle to convert exposure into durable improvement.

Research on AI and self-regulated learning has expanded quickly. Chiu (2024) proposes a classification tool to foster self-regulated learning with generative AI, grounded in principles derived from self-determination theory, arguing that AI support should enhance autonomy, competence, and reflective choice. Xia et al. (2023) found that self-regulated

learning through AI chatbots is influenced by learners' psychological need satisfaction, suggesting that AI effectiveness depends partly on how learners perceive control and support during interaction. Dizon et al. (2025) report that university EFL students use ChatGPT for self-regulated language learning, involving planning, practice, and monitoring, but also note uneven strategy quality and a need for stronger pedagogical framing.

Systematic evidence reinforces both the promise and the challenge. Guan et al. (2025) conclude that educational chatbots can assist with planning, monitoring, feedback, and reflection, which are core functions of self-regulation. However, the review also shows that many chatbot interventions remain short-term, narrowly task-bound, or insufficiently integrated with curricular routines. In language learning, this is a serious concern because strong self-regulation depends on continuity across multiple practice episodes. An agentic AI system that stores goals, prompts reflection, and adapts next steps may offer a closer fit to the cyclical nature of self-regulated language learning than reactive tools that treat each exchange as independent.

There is also growing evidence that AI can contribute to measurable language achievement when its support is sustained and pedagogically aligned. Wei (2023) reported positive effects of AI in language instruction on English learning achievement, L2 motivation, and self-regulated learning. Meta-analytic findings likewise indicate overall positive effects of AI on English language achievement, although the strength of these effects varies across instructional designs and social-emotional conditions (Xu & Wang, 2024; Wu, 2024). The remaining question is whether agentic features make a distinctive difference. If AI can proactively support planning, revision, and reflection, then self-regulated learning may become the pathway to stronger language outcomes.

D. Research Questions

Based on the foregoing literature and conceptual framing, the study addressed the following research questions:

RQ1. To what extent is agentic AI-supported ESL instruction associated with higher posttest English proficiency than reactive AI-supported ESL instruction?

RQ2. What differences emerge between the two conditions in learner engagement, self-regulated language learning, and agentic engagement?

RQ3. Which patterns of agentic AI use predict posttest English proficiency within the experimental group?

RQ4. How do students and instructors perceive the influence of agentic AI on ESL learning processes and classroom pedagogy?

III. METHODOLOGY

A. Research Design and Context

The study employed an explanatory sequential mixed-methods, quasi-experimental design. Quantitative and qualitative phases were connected so that interview data could explain patterns observed in performance scores, survey responses, and learning analytics. The intervention lasted 12 weeks and was implemented in four intact undergraduate ESL sections across two private universities in the Philippines. The participating sections followed a shared syllabus for academic communication that emphasized paragraph writing, oral presentations, vocabulary development, source-based reading, and revision strategies.

To reduce instructional variance, the two course instructors worked from a shared weekly plan, common assessment rubrics, matched task prompts, and coordinated pacing. Each instructor handled one agentic AI section and one reactive AI section. Both conditions received the same contact hours, task genres, teacher feedback schedule, and grading criteria. The core difference lay in the AI support architecture used outside and during selected in-class activities.

B. Participants

Participants were 128 undergraduate students enrolled in required ESL courses. Sixty-four students were assigned to the agentic AI condition and 64 to the reactive AI condition via intact-section membership. Students ranged from first to third year and represented intermediate to upper-intermediate English proficiency bands based on local placement records and pretest performance. Table 1 summarizes the participant profile by condition.

Pre-intervention checks showed no statistically significant differences between groups in age, English proficiency, learner engagement, self-regulated language learning, or agentic engagement (all $p > .05$). This supported the comparability of the two cohorts at baseline, although the intact-section design meant that causal claims should still be interpreted with appropriate caution.

TABLE 1
PARTICIPANT CHARACTERISTICS BY CONDITION

Characteristic	Total (N = 128)	Agentic AI (n = 64)	Reactive AI (n = 64)
n	128	64	64
Age, M (SD)	19.46 (1.14)	19.38 (1.10)	19.54 (1.18)
Female, n (%)	79 (61.7)	40 (62.5)	39 (60.9)
Male, n (%)	44 (34.4)	21 (32.8)	23 (35.9)
Non-binary/Prefer not to say, n (%)	5 (3.9)	3 (4.7)	2 (3.1)
B1, n (%)	73 (57.0)	37 (57.8)	36 (56.2)
B2, n (%)	55 (43.0)	27 (42.2)	28 (43.8)
First-year, n (%)	72 (56.2)	37 (57.8)	35 (54.7)
Second-year, n (%)	41 (32.0)	22 (34.4)	19 (29.7)
Third-year, n (%)	15 (11.7)	5 (7.8)	10 (15.6)

Note. Percentages are column percentages.

C. Instructional Intervention

Students in the experimental condition used an Agentic English Coach (AEC) embedded within the institutional learning management system. The AEC was configured as a bounded instructional agent rather than an unrestricted conversational tool. At the start of each week, the system asked students to set a target related to the week's language outcome, such as producing a thesis-driven paragraph, improving oral fluency for a short presentation, or mastering discipline-specific vocabulary. It then decomposed the target into a sequence of tasks, scheduled suggested checkpoints, and recommended resources from a teacher-approved repository. During the week, the system monitored task completion, issued reminders, generated focused practice prompts, and requested brief reflections after drafts or speaking attempts. When the system detected repeated uncertainty, contradictory outputs, or low-confidence interactions, it prompted the learner to consult the teacher and flagged the issue in the instructor dashboard.

To clarify the implementation boundaries, the AEC was configured through teacher-authored weekly workflow templates rather than an open-ended autonomous agent. The templates defined the course outcome, task sequence, resource bank, reminder schedule, reflection prompt, and escalation trigger for each week. The learning management system-integrated AI service adapted prompts, practice items, and feedback focus within those constraints; it did not assign grades, revise the syllabus, or make placement decisions. The study therefore evaluated a bounded process-scaffolding configuration rather than a comparison among model vendors or model architectures.

The reactive AI condition used the same learning management system and had access to a standard AI chat interface for idea generation, clarification, feedback requests, and language practice. However, the reactive interface lacked cross-session memory, automatic task planning, a progress dashboard, a reminder function, and proactive reflection prompts. Students could obtain help when they initiated it, but the system did not manage or sequence the learning activity. This contrast allowed the study to compare two forms of AI-supported pedagogy that differed primarily in instructional agency and process orchestration rather than in access to AI itself.

Figure 1 presents the instructional cycle used in the agentic AI condition. The figure illustrates how weekly goal negotiation, task decomposition, guided practice, monitoring, reflection, and teacher oversight formed a recurring loop. This design treated AI not as a substitute for instruction, but as a process manager that helped students sustain productive language learning routines between teacher contact points.

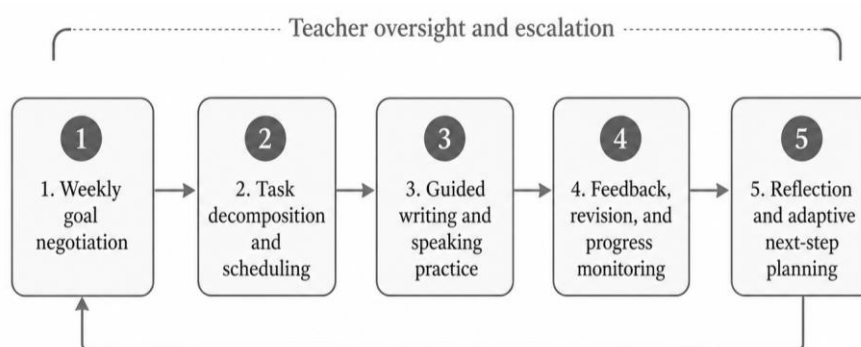


Figure 1. Weekly Instructional Cycle in the Agentic AI Condition

D. Instruments

English proficiency was measured through an integrated performance assessment administered in Weeks 1 and 12. The assessment combined a 250-word argumentative writing task and an individual speaking task lasting approximately four minutes. Writing and speaking were scored using analytic rubrics that covered organization, language control, vocabulary use, task fulfillment, and fluency. Scores were converted to a composite scale from 0 to 100. Two trained raters independently scored all performances. Interrater agreement, calculated through intraclass correlation, was .90 at pretest.

and .92 at posttest.

Learner engagement was measured with a 12-item scale adapted from the multidimensional engagement framework of Fredricks et al. (2004) and the agentic engagement work of Reeve and Tseng (2011). The scale covered behavioral participation, emotional involvement, cognitive effort, and agentic contribution. Responses were collected on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Internal consistency was acceptable in the present sample ($\alpha = .89$ at pretest and $.91$ at posttest).

Self-regulated language learning was measured with a 12-item scale aligned with the forethought, performance, and reflection phases described by Zimmerman (2002) and summarized by Panadero (2017). Items targeted goal setting, planning, monitoring, strategic action, help seeking, and reflection. The same 5-point response format was used. Reliability was acceptable ($\alpha = .87$ at pretest and $.90$ at posttest).

Agentic engagement was examined both as a subdimension of learner engagement and as a focal outcome in its own right, through four targeted items: expressing preferences, seeking clarification, requesting challenge adjustments, and reshaping the sequence of learning activities. These items were analyzed separately to capture the specific learner behaviors most directly related to the agentic AI intervention. Reliability was acceptable ($\alpha = .84$ at pretest and $.87$ at posttest).

System log variables were collected for students in the experimental condition. These included plan completion percentage, reflection completion percentage, number of revision cycles, number of speaking practice turns, and number of teacher escalation flags. Finally, semi-structured interviews were conducted with 12 students selected to reflect variation in proficiency gain and platform usage, along with the two course instructors and two program coordinators who observed the implementation.

E. Procedure

Week 1 was used for orientation, consent, pretesting, and platform familiarization. In Weeks 2 through 11, students completed weekly instructional cycles tied to the course syllabus. Both conditions worked on the same genres and language targets, including paragraph development, oral rehearsal, vocabulary application, and source-based response writing. In the experimental condition, students began by confirming weekly goals in the AEC, completed staged tasks, and submitted short reflections after major activities. In the reactive condition, students accessed the AI tool as needed but did not follow any system-generated learning path.

Teachers reviewed classroom outputs in both conditions, addressed recurring language issues during scheduled lessons, and monitored academic integrity. In the experimental condition, teacher dashboards displayed incomplete plans, repeated error patterns, and escalation flags. Teachers used these signals to offer clarification and brief corrective feedback during the following class meeting. Posttests were administered in Week 12, after which interviews were conducted and transcribed for analysis.

F. Data Analysis

Independent-samples t-tests were used to assess baseline equivalence. To address RQ1 and RQ2, analyses of covariance (ANCOVAs) were conducted for posttest English proficiency, learner engagement, self-regulated language learning, and agentic engagement, using the corresponding pretest scores as covariates. Effect sizes were reported as partial eta squared. To address RQ3, multiple regression was performed within the experimental group to determine whether platform-use variables predicted posttest proficiency, controlling for pretest proficiency.

Qualitative data were analyzed through iterative thematic coding. After initial open coding, categories were compared across student and teacher transcripts to identify recurring patterns related to pacing, practice quality, feedback use, autonomy, and teacher role. The qualitative results were then used to explain and elaborate on the quantitative findings, especially the strong gains observed in self-regulated learning and agentic engagement.

IV. RESULTS

A. Baseline Equivalence and Descriptive Statistics

Table 2 presents the descriptive statistics for all major outcomes. The two groups started from closely comparable baselines. The agentic AI group increased from a pretest mean of 59.36 (SD = 6.73) to a posttest mean of 72.59 (SD = 6.93), while the reactive AI group increased from 61.40 (SD = 6.68) to 69.18 (SD = 4.99). A similar pattern appeared in the self-report measures. The agentic AI group showed larger gains in learner engagement (mean gain = 0.67 vs. 0.25), self-regulated language learning (mean gain = 0.84 vs. 0.29), and agentic engagement (mean gain = 0.88 vs. 0.32).

TABLE 2
DESCRIPTIVE STATISTICS FOR THE MAJOR OUTCOMES BY CONDITION

Outcome	Agentic AI pretest, M (SD)	Agentic AI posttest, M (SD)	Reactive AI pretest, M (SD)	Reactive AI posttest, M (SD)
English proficiency	59.36 (6.73)	72.59 (6.93)	61.40 (6.68)	69.18 (4.99)
Learner engagement	3.20 (0.39)	3.87 (0.25)	3.18 (0.35)	3.43 (0.27)
Self-regulated language learning	3.08 (0.35)	3.91 (0.27)	3.10 (0.40)	3.39 (0.24)
Agentic engagement	3.09 (0.46)	3.97 (0.27)	3.01 (0.45)	3.33 (0.25)

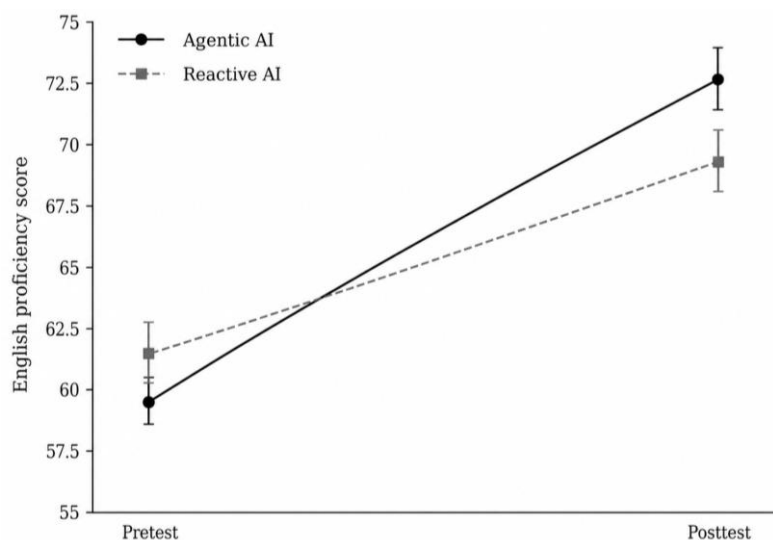


Figure 2. Pretest-to-Posttest Change in English Proficiency by Condition
Note. Error bars represent standard errors.

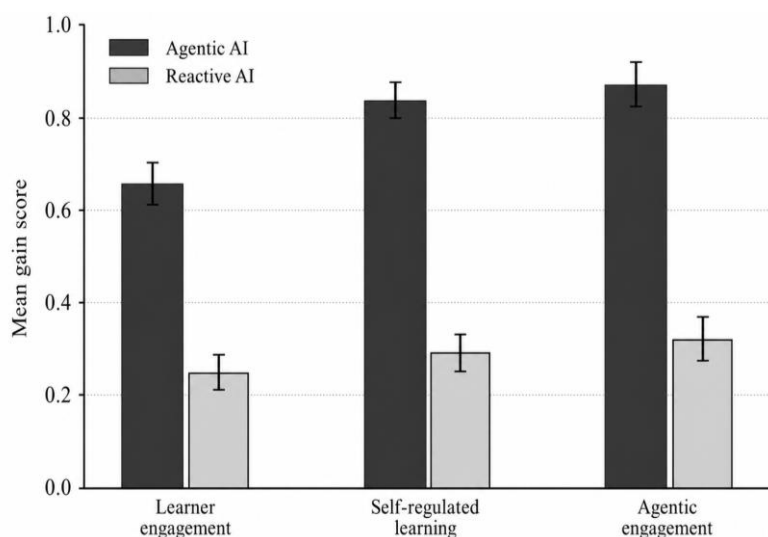


Figure 3. Mean Gain Scores for Learner Engagement, Self-Regulated Learning, and Agentic Engagement
Note. Error bars represent standard errors.

These descriptive patterns suggest that both conditions benefited from AI-supported learning, but the gains were consistently stronger when AI was configured to manage learning processes over time. Figure 2 visualizes the pretest-to-posttest change in English proficiency by condition, while Figure 3 summarizes mean gains across the three engagement-related outcomes.

B. Effects on English Proficiency, Engagement, and Self-Regulated Learning

The ANCOVA results in Table 3 showed a statistically significant effect of condition on posttest English proficiency after controlling for pretest scores, $B = 4.57$, $SE = 0.84$, $t = 5.42$, $p < .001$, partial eta squared = .19. The adjusted posttest mean for the agentic AI group was 73.16, compared with 68.60 for the reactive AI group. The raw-gain difference of approximately 5.45 points indicates a meaningful instructional advantage for the agentic AI condition over the 12-week period.

TABLE 3
ANCOVA RESULTS FOR POSTTEST OUTCOMES CONTROLLING FOR CORRESPONDING PRETEST SCORES

Outcome	Adjusted M Agentic	Adjusted M Reactive	B	SE	t	p	Partial η^2
English proficiency	73.16	68.60	4.57	0.84	5.42	< .001	0.19
Learner engagement	3.86	3.44	0.43	0.03	12.88	< .001	0.57
Self-regulated language learning	3.92	3.39	0.53	0.03	15.97	< .001	0.67
Agentic engagement	3.96	3.34	0.62	0.04	15.57	< .001	0.66

Note. Each model used the matching pretest measure as a covariate.

The effects on process-oriented outcomes were even stronger. For learner engagement, the condition effect was

significant, $B = 0.43$, $SE = 0.03$, $t = 12.88$, $p < .001$, partial eta squared = .57. The adjusted posttest means were 3.86 for the agentic AI group and 3.44 for the reactive AI group. For self-regulated language learning, the condition effect was also significant, $B = 0.53$, $SE = 0.03$, $t = 15.97$, $p < .001$, partial eta squared = .67, with adjusted posttest means of 3.92 and 3.39, respectively. Agentic engagement produced a similarly large effect, $B = 0.62$, $SE = 0.04$, $t = 15.57$, $p < .001$, partial eta squared = .66, yielding adjusted means of 3.96 for the agentic AI group and 3.34 for the reactive AI group.

Taken together, these results indicate that the intervention's strongest effects emerged in the learning processes that agentic AI was explicitly designed to support. The system did not simply improve endpoint performance; it altered how students initiated, sustained, and adjusted their language learning activity during the semester.

C. Learning Analytics Predictors of Posttest English Proficiency

To understand which forms of system use were associated with stronger performance within the experimental group, a multiple regression model was estimated using posttest English proficiency as the dependent variable and pretest proficiency as a control. The overall model was significant, $F(6, 57) = 37.19$, $p < .001$, and explained 79.7% of the variance in posttest proficiency (R squared = .80, adjusted R squared = .78). Table 4 presents the full regression coefficients.

TABLE 4
MULTIPLE REGRESSION PREDICTING POSTTEST ENGLISH PROFICIENCY IN THE AGENTIC AI GROUP

Predictor	B	SE	β	t	p
Pretest proficiency	0.53	0.07	0.51	8.07	< .001
Plan completion (%)	0.12	0.05	0.19	2.48	0.016
Reflection completion (%)	0.11	0.05	0.17	2.23	0.029
Revision cycles	1.29	0.44	0.23	2.95	0.005
Speaking practice turns	0.08	0.04	0.21	2.33	0.023
Teacher escalations	-0.35	0.30	-0.08	-1.16	0.249

Note. Model statistics: R squared = 0.80, adjusted R squared = 0.78, $F(6, 57) = 37.19$, $p < .001$.

Pretest proficiency remained the strongest predictor of posttest performance ($\beta = .51$, $p < .001$), but four engagement-with-process variables also made significant positive contributions. Plan completion predicted higher posttest proficiency ($B = 0.12$, $\beta = .19$, $p = .016$), as did reflection completion ($B = 0.11$, $\beta = .17$, $p = .029$), revision cycles ($B = 1.29$, $\beta = .23$, $p = .005$), and speaking practice turns ($B = 0.08$, $\beta = .21$, $p = .023$). Teacher escalations carried a negative coefficient but did not reach statistical significance ($B = -0.35$, $\beta = -.08$, $p = .249$).

This pattern is notable because it suggests that productive use of agentic AI was not captured by access alone. Students benefited most when they actually completed plans, reflected on performance, revised repeatedly, and engaged in sustained speaking rehearsal. The non-significant negative coefficient for teacher escalations may indicate that students who required more intensive intervention were also those who faced greater language difficulties, rather than that teacher support itself reduced performance.

D. Qualitative Findings

The interview data helped explain why the agentic AI condition produced stronger gains in engagement and self-regulated language learning. Four themes were identified: proactive structuring reduced uncertainty, visible progress monitoring strengthened self-regulation, iterative practice improved communicative confidence, and teacher oversight remained essential for pedagogical trust. Table 5 summarizes these themes with representative excerpts.

TABLE 5
QUALITATIVE THEMES EXPLAINING THE AGENTIC AI EFFECT

Theme	Description	Representative excerpt
Proactive structuring reduced uncertainty	Students described the agentic system as breaking large tasks into manageable weekly actions that reduced procrastination and clarified what to do next.	"The weekly steps made the task feel smaller. I did not wait for the deadline because I already knew the next action." (S4)
Visible progress monitoring strengthened self-regulation	Dashboards, completion markers, and reflection prompts helped learners compare earlier and later work and notice patterns in language weaknesses.	"The reflection box made me look at my first and last draft. I could see the same grammar problem coming back." (S9)
Iterative practice improved communicative confidence	Repeated speaking turns and revision loops encouraged rehearsal, experimentation, and more confident oral and written production.	"For speaking, it kept asking me to try again with better transitions. After a few rounds, I sounded less hesitant." (S2)
Teacher oversight remained essential for pedagogical trust	Students valued the system most when teachers verified feedback, clarified inaccuracies, and intervened when the platform signaled repeated difficulty.	"My job became checking the quality of the path students were following, not only the final product they submitted." (T1)

Students in the agentic AI group repeatedly described the AEC as a system that kept them moving. Rather than waiting until deadlines were near, they reported being nudged into smaller weekly actions. Several participants noted that the platform's task sequencing made large assignments feel more manageable. This perception aligns with the quantitative gains in behavioral and agentic engagement because students felt less dependent on moment-to-moment willpower and

more guided by a visible process.

A second theme concerned monitoring and reflection. Students explained that the dashboard, completion markers, and reflection prompts made it easier to notice weak areas across writing and speaking tasks. Instead of stopping after receiving a generated answer, they compared earlier and later versions of their work, which helped them identify patterns in grammar, cohesion, and vocabulary use. This theme provides a strong qualitative explanation for the sizable differences observed in self-regulated language learning.

Interviewees also emphasized the value of iterative rehearsal, especially for speaking. The system's repeated invitations to try again, varied expressions, and rehearsal of transitions appeared to reduce hesitation and increase practice volume. At the same time, both students and teachers stressed that human oversight was necessary when AI feedback became too generic, overly confident, or inconsistent with classroom expectations. Instructors described their role as shifting from constant explanation toward orchestration, verification, and timely intervention. This suggests that the pedagogical strength of agentic AI lies in coordinated support rather than in replacing teacher judgment.

V. DISCUSSION

The findings suggest that agentic AI can contribute to ESL pedagogy not simply by improving the convenience of question answering, but by structuring learning activity across time. When AI was designed as a bounded process scaffolding, students achieved higher posttest proficiency and reported stronger engagement, self-regulated language learning, and agentic participation than peers who used a reactive interface. The pattern supports the argument that the pedagogical value of AI depends less on the quality of isolated responses than on its alignment with the cyclical nature of language learning.

The proficiency results extend previous evidence that AI can support English learning achievement (Wei, 2023; Xu & Wang, 2024; Wu, 2024). However, the present study adds a more specific interpretation. The agentic AI advantage did not appear to stem merely from having better language generation available. Both groups had AI access. The difference was that one group used a system that translated course outcomes into sequenced actions, monitored completion, prompted reflection, and encouraged return visits. This helps explain why the raw proficiency gain was larger in the agentic AI condition despite comparable baseline levels and shared instruction.

The much stronger effects for learner engagement, self-regulated learning, and agentic engagement are especially important. These are precisely the processes that agentic AI was intended to influence. The results therefore align well with the work of Reeve and Tseng (2011), who describe agentic engagement as learners' constructive contributions to the flow of instruction. In the present intervention, students were repeatedly invited to choose goals, request more challenge, reflect on obstacles, and adjust next steps. The technology effectively created more opportunities for learners to act on the instruction rather than simply receive it. Reeve and Shin (2020) argue that classrooms can be structured to support this kind of agentic contribution, and the current results suggest that well-bounded AI systems may help operationalize that principle.

The study also reinforces self-regulated learning as a key mechanism. Zimmerman's (2002) model emphasizes that strong performance arises from forethought, performance monitoring, and self-reflection. The intervention mapped closely onto those phases. Students planned weekly goals, monitored completion, received targeted prompts during practice, and reflected after drafts and speaking turns. The regression findings are consistent with this interpretation because plan completion, reflection completion, revision cycles, and speaking practice turns all predicted posttest proficiency within the experimental group. These are not simply indicators of exposure. They are indicators of sustained, strategic participation in the learning process.

This process-oriented interpretation resonates with recent literature on AI and language learning. Chiu (2024), Xia et al. (2023), and Dizon et al. (2025) all highlight the importance of autonomy-supportive scaffolding, psychological need satisfaction, and guided strategy use when learners interact with generative AI. Guan et al. (2025) similarly argue that chatbot support is most educationally powerful when it assists planning, monitoring, and reflection rather than isolated content retrieval. The present study builds on these insights by showing that when such functions are organized into a persistent agentic system, they can produce not only perceived benefits but measurable classroom differences.

The qualitative findings further illuminate why these effects emerged. Students repeatedly described the system as reducing uncertainty by converting broad language tasks into smaller actions. That matters in ESL pedagogy because many learners know what they are expected to produce but are less certain about how to distribute effort across time. Agentic AI appears to have served as a bridge between curricular outcomes and daily learning behavior. This supports Liu et al.'s (2025) suggestion that AI-mediated English learning becomes more meaningful when learners adopt agentic digital practices rather than passive consumption.

The interviews also reveal a crucial boundary condition: teacher oversight remained indispensable. Students trusted the system most when its suggestions were tied to teacher-approved materials and when escalation pathways existed for confusing or inaccurate feedback. Instructors did not disappear from the pedagogy. Instead, their role shifted toward orchestration, verification, and targeted intervention. This echoes Molenaar's (2022) concept of hybrid human-AI regulation and offers a practical design principle for ESL programs. AI should manage recurrence and routine, while teachers should remain responsible for interpretation, quality control, and ethical framing.

Several implications for language teaching follow from these findings. First, AI in ESL programs should be designed

as a bounded coaching layer for goal setting, pacing, rehearsal, and reflection, not merely as a source of generated text. Second, teachers should make process indicators visible, because completion data, revision frequency, and reflection logs can become actionable forms of formative information when linked to classroom feedback. Third, teacher oversight should remain integral. Escalation rules, approved resource banks, and scheduled verification help ensure that AI support stays aligned with course outcomes and does not displace human judgment. Finally, assessment design should reward revision, rehearsal, and strategic engagement alongside final performance. In this sense, agentic AI is most pedagogically useful when it strengthens teacher-led language instruction rather than automating it.

The study has several limitations. The quasi-experimental design relied on intact sections, so random assignment was not possible and causal claims should remain cautious. The intervention lasted one semester, which limits conclusions about longer-term retention or transfer. The proficiency measure combined writing and speaking into a single composite score, and the engagement measures were self-reported. In addition, the study examined one bounded agentic AI configuration in one national context. Different course levels, institutional AI policies, or system designs may produce different patterns. Future research could use randomized or longitudinal designs, separate writing and speaking outcomes, and compare specific agentic features within teacher-supervised language courses.

VI. CONCLUSION

This study shows that the pedagogical significance of agentic AI in ESL instruction lies in its capacity to organize language learning as a sustained process rather than a series of isolated prompts. Across a 12-week intervention, students who used a bounded agentic system achieved higher posttest English proficiency and reported stronger learner engagement, self-regulated language learning, and agentic engagement than students who used a reactive AI interface. Learning analytics showed that gains were associated with concrete process behaviors, especially plan completion, reflection, repeated revision, and speaking practice. Interview data indicated that these advantages emerged because the system reduced task uncertainty, made progress more visible, and supported iterative rehearsal, while teacher oversight remained essential for accuracy, trust, and curricular alignment. The central implication is clear: in language education, AI is most valuable when it helps learners plan, persist, revise, and reflect within teacher-led instruction. For ESL pedagogy, agentic AI is best understood not as a replacement for teachers, but as a bounded process scaffolding that makes self-regulated learning more actionable in everyday coursework.

APPENDIX. SURVEY ITEMS USED IN THE STUDY

Response scale for all items: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree.

Learner Engagement Scale

1. I completed the English learning tasks assigned for each week on time.
2. I stayed focused during AI-supported English activities even when they were challenging.
3. I participated actively in class and online tasks related to the course.
4. I felt interested in the English tasks I worked on during the course.
5. I enjoyed using course activities to improve my English.
6. I tried to understand ideas deeply instead of finishing tasks quickly.
7. I connected new English content with what I had already learned.
8. I looked for better ways to improve my work when feedback showed weaknesses.
9. I asked for clarification when I did not understand a task or feedback.
10. I suggested adjustments or requested extra practice when I needed it.
11. I used the AI tool to shape my learning process, not only to get answers.
12. I took initiative in deciding how to approach my weekly English goals.

Self-Regulated Language Learning Scale

1. At the start of the week, I set clear goals for what I wanted to improve in English.
2. I made a realistic plan for finishing my English tasks.
3. I broke large English tasks into smaller steps.
4. I checked my progress while working instead of waiting until the end.
5. I noticed quickly when I did not understand a language point.
6. I changed strategies when my first approach was not working.
7. I revised my writing after receiving feedback from the teacher or AI tool.
8. I practiced speaking more than once before submitting or performing.
9. I sought help when I could not solve a language problem by myself.
10. I compared my final work with earlier drafts or attempts.
11. I reflected on which study habits helped me improve my English.
12. I used what I learned from one task to plan for the next task.

REFERENCES

- [1] Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development*, 72, 2401-2416. <https://doi.org/10.1007/s11423-024-10366-w>
- [2] Dizon, G., Gold, J., & Barnes, R. (2025). ChatGPT for self-regulated language learning: University English as a foreign language students' practices and perceptions. *Digital Applied Linguistics*, 3, 102510. <https://doi.org/10.29140/dal.v3.102510>
- [3] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- [4] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- [5] Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, 30(4), 4493-4518. <https://doi.org/10.1007/s10639-024-12881-y>
- [6] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [7] Liu, G. L., Lee, J. S., & Zhao, X. (2025). Critical digital literacies, agentic practices, and AI-mediated informal digital learning of English. *System*, 134, 103797. <https://doi.org/10.1016/j.system.2025.103797>
- [8] Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence*, 3, 100070. <https://doi.org/10.1016/j.caeai.2022.100070>
- [9] 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>
- [10] Plaat, A., van Duijn, M., van Stein, N., Preuss, M., van der Putten, P., & Batenburg, K. J. (2025). Agentic large language models, a survey. *Journal of Artificial Intelligence Research*, 84, Article 29. <https://doi.org/10.1613/jair.1.18675>
- [11] Reeve, J., & Shin, S. H. (2020). How teachers can support students' agentic engagement. *Theory Into Practice*, 59(2), 150-161. <https://doi.org/10.1080/00405841.2019.1702451>
- [12] Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- [13] Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System*, 127, 103528. <https://doi.org/10.1016/j.system.2024.103528>
- [14] Wang, Y., & Xue, L. (2024). Using AI-driven chatbots to foster Chinese EFL students' academic engagement: An intervention study. *Computers in Human Behavior*, 159, 108353. <https://doi.org/10.1016/j.chb.2024.108353>
- [15] Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14, 1261955. <https://doi.org/10.3389/fpsyg.2023.1261955>
- [16] Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System*, 126, 103498. <https://doi.org/10.1016/j.system.2024.103498>
- [17] Xia, Q., Chiu, T. K. F., Chai, C. S., & Xie, K. (2023). The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot. *British Journal of Educational Technology*, 54(4), 967-986. <https://doi.org/10.1111/bjet.13305>
- [18] Xu, T., & Wang, H. (2024). The effectiveness of artificial intelligence on English language learning achievement. *System*, 125, 103428. <https://doi.org/10.1016/j.system.2024.103428>
- [19] Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, 103351. <https://doi.org/10.1016/j.system.2024.103351>
- [20] Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, 127, 103501. <https://doi.org/10.1016/j.system.2024.103501>
- [21] Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/S15430421TIP4102_2

Roland A. Niez was born in Biliran Province, Philippines. He earned a Doctor of Arts in Language Teaching from Leyte Normal University in 2024 and a Doctor of Education from Biliran Province State University. He is Vice-President for Research, Innovation, and Societal Impact at Biliran Province State University. Dr. Niez's research interests include second language acquisition, multilingual education, culturally responsive pedagogy, and innovative ESL instruction. ORCID iD: 0009-0006-2591-1208.

Joedel O. Peñaranda is an English and Humanities educator based in Melbourne, Australia. He holds a Master of Arts in Education from the University of San Carlos and a Master in Public Management from the Development Academy of the Philippines. His work focuses on research-informed pedagogy, curriculum design, assessment, debate and journalism coaching, and educational innovation. ORCID iD: 0000-0002-7762-4431. Email: jpenaranda@bipsu.edu.ph

Maribel N. Zipagan was born in Biliran Province, Philippines, on November 17, 1980. She earned a Ph.D. in English Language Education from Sejong University in 2019 and an M.A. in TESOL from Hankuk University of Foreign Studies in 2012. She is Vice President for Students, Internationalization, and Strategic Partnerships and Associate Professor of English at Biliran Province State University. Dr. Zipagan's research interests include language learning, speaking assessment, and writing assessment. ORCID iD: 0000-

0003-0776-6378.

John Cavin M. Sabonsolin was born in Naval, Biliran Province, Philippines, on February 2, 1995. He earned a Doctor of Arts in Literature and Communication from Cebu Normal University in 2023, a Master of Media Studies in 2018, and a Bachelor of Arts in Mass Communication in 2015. He is Program Chair of the Department of Communication at Biliran Province State University. Dr. Sabonsolin's research interests include communication, media studies, literature, language studies, and media literacy. ORCID iD: 0000-0002-4638-2484.

Cristie Ann L. Jaca earned degrees in English language teaching, education, early childhood education, and development education from the University of San Carlos and Cebu Technological University. She is affiliated with Cebu Technological University, Cebu City, Philippines. Dr. Jaca's scholarly interests include English language teaching, early childhood education, teacher learning, environmental education, and meaningful innovation in education. Email: cristieann.jaca@ctu.edu.ph

Noel P. Tancinco was born in Tucdao, Biliran Province, Philippines. He earned a Ph.D. in Educational Programs Management from Eastern Visayas State University and is pursuing a second Ph.D. in Filipino at the University of Mindanao. He is Associate Professor V and Director of Academic-Related Services at Biliran Province State University. Dr. Tancinco's research interests include Filipino linguistics, pedagogy, educational leadership, management systems, innovative teaching methodologies, and language education. Email: noel.tancinco@bipsu.edu.ph

Jelyn M. Magno was born in Biliran, Philippines, on August 20, 1971. She earned a Doctor of Education degree in Filipino Language Teaching from Cebu Normal University in 2000. She is Professor I and Program Chair of the Bachelor of Secondary Education program at Biliran Province State University-Biliran Campus. Dr. Magno's research interests include discourse analysis, Filipino language teaching, and creative language education strategies. ORCID iD: 0000-0001-8692-4287.

Mary Joy B. Gelizon was born in Biliran, Biliran Province, Philippines, on January 7, 1986. She earned a Doctor of Education in Early Childhood Education from Cebu Normal University in 2021 and previously served as a Department of Education teacher. She is the BEED OJT and Guidance Coordinator at Biliran Province State University-Biliran Campus. Dr. Gelizon's research interests include linguistics, pedagogy, inclusive education, early childhood education, and age-appropriate instructional innovation. ORCID iD: 0000-0003-0767-7359.

Jed Priam Cerera is affiliated with Biliran Province State University, Naval, Biliran, Philippines. His work in this study relates to AI-supported ESL instruction, learner engagement, self-regulated language learning, and educational technology. Email: jedpriam@proton.me