Impact of Generative AI on Self-Regulated Learning and Cognitive Offloading

Maria Rizalie Lindo [1,*], Cherry Ann Cutad [2]

[1] Basic Education Department-Senior High School Unit, University of the Immaculate Conception, Davao, Philippines
[2] College of Development Management, University of Southeastern Philippines, Davao, Philippines; cherry.cutad@usep.edu.ph (C.A.C.)

* Correspondence: mlindo@uic.edu.ph

Abstract

Generative artificial intelligence (AI) is rapidly reshaping how senior high school students study, yet its implications for self-regulated learning and learning durability remain poorly resolved in authentic classrooms. This study examined whether a structured generative AI approach strengthens self-regulated learning while increasing cognitive offloading, and whether offloading predicts weaker delayed retention. Using a cluster randomised pretest and posttest design, intact class sections from one public and one private senior high school in Davao City, Philippines were assigned either to structured generative AI use guided by plan, monitor, and evaluate prompts or to non-AI study support. Self-regulated learning was measured through adapted MSLQ strategy scales, cognitive offloading through a brief offloading tendency scale, and achievement through unit quizzes including a two-week delayed retention assessment. Mixed-effects models accounted for section-level clustering and adjusted for baseline scores and access indicators. Structured generative AI use was associated with higher posttest self-regulated learning and substantially higher cognitive offloading. Offloading was negatively associated with delayed retention after adjustment, while the direct condition effect on delayed retention was not reliable. These findings underscore the need to integrate generative AI as a scaffold for verification and metacognitive control rather than as a substitute for reasoning.

1. Introduction

Generative artificial intelligence (AI) has become embedded in everyday study practices, yet its educational consequences remain incompletely specified. In particular, the field lacks sustained classroom evidence on two constructs that are theoretically proximal to long-term learning: self-regulated learning and cognitive offloading. Self-regulated learning concerns how learners plan, monitor, and control cognition, motivation, and behaviour in pursuit of academic goals, and it is widely treated as a mechanism through which instructional supports translate into durable performance gains [1,2]. Contemporary syntheses emphasise that self-regulation is neither a trait nor a slogan, but a set of context sensitive processes that can be strengthened or weakened by tool environments and assessment incentives [3,4].

Generative AI may plausibly support self-regulated learning when it is used to scaffold planning, iterative refinement, and evaluation. It may equally undermine self-regulation when it compresses the feedback loop into rapid answer provision, reducing the need for self-monitoring and strategic adjustment. The second mechanism, cognitive offloading, sharpens this concern. Cognitive offloading refers to the use of external tools or actions to reduce internal processing demands, and the construct has matured from a descriptive label into a coherent account of how people allocate effort across internal and external resources [5]. Offloading decisions are linked to metacognitive judgements about memory and confidence, which makes offloading inseparable from self-regulation rather than merely a by-product of convenience [6,7]. Digital search environments already reshape what people choose to remember and what they delegate to external systems, suggesting that the distribution of cognitive labour adapts to tool availability [8-12]. Generative AI extends this logic beyond storage and retrieval by generating explanations, arguments, and solution steps, thereby shifting not only what is remembered, but what is practised.

The empirical literature on generative AI in education is expanding rapidly. Evidence syntheses suggest that ChatGPT can improve learning performance on average, but effects vary substantially by task type, instructional framing, and the role assigned to the tool, which implies that pedagogy is a decisive moderator rather than a minor implementation detail [13]. At the same time, recent conceptual work has described a cognitive paradox, where AI can enhance engagement under scaffolding conditions and erode engagement under substitution conditions [14]. Concerns about durability are also consistent with established findings from learning science. Techniques that reliably support retention and transfer, such as retrieval practice and spacing, depend on effortful processing, and those conditions may be attenuated when learners externalise core generative work [9,10].

Broader educational guidance on large language models therefore tends to converge on a pragmatic position: the tools are neither uniformly beneficial nor uniformly harmful, and outcomes hinge on how learners are required to check, explain, and revise their own thinking [11,12].

Secondary school contexts remain underrepresented in this evidence base, particularly outside North America and Europe. In the Philippines, senior high school learners operate under heterogeneous conditions that include variability in device access, internet stability, class size, and supplementary academic supports. Differences between public and private schools may therefore shape both exposure and the functional role of generative AI, with implications for equity and learning integrity. Against this background, the present study evaluates the association of structured generative AI use with self-regulated learning and cognitive offloading among senior high school students in Davao City, Philippines, sampling one public and one private school. The study further examines whether cognitive offloading is associated with delayed retention after adjustment for baseline achievement and self-regulation.

2. Methodology

2.1 Design and Setting

The study used a cluster randomised pretest and posttest design in which intact class sections served as the unit of randomisation and analysis clustering. The setting comprised one public and one private senior high school in Davao City, Philippines. The intervention period spanned six instructional weeks within a single academic term, with baseline assessment conducted in the week immediately preceding implementation and follow up assessment conducted two weeks after posttest.

2.2 Participants

Participants were Grade 11 and Grade 12 students enrolled in participating sections. Eligibility required enrolment in a selected section, capacity to provide assent, and guardian consent for minors. Students with incomplete consent documentation were not enrolled in study procedures, and they continued normal class activities without penalty. Recruitment was conducted through school coordination meetings and class level information sessions. To minimise coercion, study staff, not classroom teachers, administered consent and assent procedures. Participation was explicitly separated from grading, and teachers did not access individual survey responses.

Sample size planning used the cluster design with conservative assumptions regarding intraclass correlation and anticipated attrition. The target number of sections per condition within each school was set to support stable estimation of section level random effects and to avoid inflated Type I error associated with very small cluster counts. The intended sample provided precision to detect small to moderate standardised differences in self-regulated learning and offloading outcomes under plausible intraclass correlation values commonly observed in classroom research.

2.3 Conditions

Clusters were defined as class sections scheduled for the same subject unit during the implementation window. Within each school, participating sections were listed and assigned a unique identifier. Randomisation was performed within school to balance context and teacher factors across conditions. A reproducible randomisation script was used to generate allocation sequences using a computer based random number generator with a fixed seed recorded in a locked study log. Allocation concealment was maintained until baseline data collection was completed. After baseline assessment, condition assignments were disclosed to the implementing teachers through written instructions. The target design included an equal number of sections per condition within each school. Any deviations from target allocation were documented with dated rationale.

A single, prespecified instructional unit was selected per participating subject, with the same learning objectives, lesson sequence, and assessment blueprints used across conditions. Teachers followed a common unit guide developed collaboratively with school representatives. The guide specified the number of class meetings, required readings, practice activities, and formative checks. Time on task was standardised by design, with both conditions completing the same in class and assigned study tasks, differing only in the permitted study support tool.

Students assigned to the structured generative AI condition used a generative AI system during designated task windows for the unit. To ensure reproducibility, the AI platform, access mode, and model name were fixed for the study period and recorded in the protocol along with the date range of use. Students received an orientation session that operationalised acceptable use as explanatory scaffolding rather than answer substitution. During each task window, students were required to use a standard prompt scaffold aligned with a plan, monitor, and evaluate cycle. The scaffold was provided verbatim in a one-page handout and embedded into the learning management platform to prevent drift across teachers and sections. Students submitted a short process log with each task consisting of their goal statement, the specific question they posed to the AI, the verification step used to check the output against class materials, and a brief reflection describing at least one correction or revision they made after verification. Teachers were instructed to grade only completion of the process log, not the content of the AI output, to reduce incentives for strategic overreporting.

Students assigned to the comparison condition completed the same tasks within the same time windows using teacher-provided materials and customary supports such as textbooks, class notes, and peer discussion. The study protocol explicitly prohibited generative AI use during the defined task windows for this condition. To reduce contamination, students were asked to keep devices closed during in class task periods except when required for non-AI activities specified in the unit guide. Outside the defined task windows, students' personal tool use was not restricted, but it was measured and treated as a covariate in sensitivity analyses.

Implementation fidelity was assessed through three complementary sources. Teachers completed brief weekly checklists documenting whether planned lessons and task windows occurred as scheduled, whether the standard instructions were delivered, and whether any deviations occurred. Students completed short adherence items after each task window indicating whether they used the permitted supports and whether any generative AI use occurred outside the assigned condition during the task. Process logs in the intervention condition served as behavioural evidence of exposure and were audited by research

Table 1: Participant characteristics by school and condition.

Characteristic	Private GenAI (n = 84)	Private Comparison (n = 78)	Public GenAI (n = 86)	Public Comparison (n = 76)
Age in years, M (SD)	17.20 (0.70)	17.10 (0.80)	17.00 (0.80)	17.10 (0.70)
Grade 12, %	46.40	44.90	43.00	42.10
Female, %	57.10	59.00	55.80	56.60
Stable home internet, %	80.90	79.50	52.30	50.00
Personal device access, %	92.90	91.00	74.40	72.40
Baseline general average (0 to 100), M (SD)	86.30 (5.80)	86.10 (6.10)	83.20 (6.40)	83.70 (6.20)

staff using a prespecified rubric that checked presence of required components, not quality of writing. Fidelity data were summarised at the section level and incorporated into per protocol analyses.

2.4 Measures

Self-regulated learning was measured using adapted learning strategy scales from the Motivated Strategies for Learning Questionnaire manual, with emphasis on metacognitive self-regulation, effort regulation, and time and study environment management [1]. Adaptation involved minor wording adjustments for senior high school comprehension while preserving construct meaning. A small pilot in a non-participating section was used to confirm item clarity and response variance. Items were scored on a fixed Likert scale with standard reverse coding applied where required. Composite scores were computed as means when at least 80 percent of items were present; otherwise, the scale score was treated as missing.

Cognitive offloading tendency was assessed using a brief scale aligned with established definitions of offloading and evidence on metacognitive drivers of offloading decisions [5-7]. Items were written to reflect common academic behaviours in the study context, including reliance on external tools to generate intermediate steps, storage of externally generated solutions, and delegation of memory demands during study. Items used a consistent response format and were aggregated by mean scoring with the same completeness rule used for self-regulated learning.

Achievement was measured with a unit aligned quiz at baseline and posttest and a delayed retention quiz administered approximately two weeks after posttest. The quiz blueprint was standardised across sections and schools and was constructed to sample both factual knowledge and applied understanding. Parallel forms were used when feasible to reduce test familiarity effects, with item difficulty and discrimination reviewed during instrument development. All quizzes were scored using an answer key with double scoring of a random subset to verify scoring accuracy.

2.5 Analysis

All survey data were collected using a secure digital form with forced response options for key fields to reduce entry errors. Paper based backups were used only if connectivity was disrupted, with subsequent double data entry. A data dictionary was created prior to analysis, defining variables, coding schemes, and derived scores. De identified datasets were stored on an encrypted drive with role-based access. An analysis log recorded all cleaning decisions, exclusions, and code revisions with timestamps to ensure auditability.

Primary inference used linear mixed effects models to account for clustering of students within class sections. For each outcome, the posttest score was modelled as a function of condition, school type, the corresponding baseline score, and prespecified covariates

capturing access stability and out of window AI use. A random intercept for section was included in all models. The condition by school interaction was estimated to explore contextual moderation. Continuous outcomes were examined for distributional assumptions and influential points, with robust standard errors used if residual diagnostics suggested heteroscedasticity. Effect sizes were reported as standardised mean differences based on model estimated marginal means for scale outcomes and as adjusted point differences for quiz outcomes. Missing data were handled via maximum likelihood under a missing at random framework, with sensitivity analyses comparing baseline characteristics of complete and incomplete cases. A per protocol analysis was additionally conducted based on prespecified adherence thresholds derived from fidelity indicators.

3. Results

3.1 Participant Characteristics

Table 1 summarises participant characteristics by school and condition. Group distributions were broadly comparable on age, grade level, and gender composition. Given contextual differences, stable home internet and personal device access were higher in the private school than in the public school.

3.2 Reliability of Measures

Internal consistency was acceptable for the composite self-regulated learning score and for cognitive offloading tendency. Omega and alpha for the self-regulated learning composite were 0.84 and 0.82, respectively. Omega and alpha for cognitive offloading tendency were 0.81 and 0.79, respectively.

3.3 Descriptive Outcomes

Table 2 presents descriptive statistics for self-regulated learning, its component strategy scales, cognitive offloading tendency, and achievement outcomes. At baseline, group means were similar for self-regulated learning and offloading. At posttest, the generative AI condition showed higher self-regulated learning and higher offloading tendency. Immediate posttest quiz performance was similar across conditions, while delayed retention was lower in the generative AI condition.

3.4 Mixed-Effects Models

Mixed-effects modelling was used to account for clustering within class sections and to adjust for baseline scores and access indicators. Table 3 reports the model predicting posttest self-regulated learning. After adjustment, the generative AI condition was associated with higher posttest self-regulated learning. Stable internet access was also positively associated with posttest self-regulation. The condition by school interaction was not statistically significant.

Table 2: Descriptive statistics for outcomes by condition.

Outcome	Time	GenAI, M (SD)	Comparison, M (SD)
Self-regulated learning composite (1 to 7)	Baseline	4.19 (0.68)	4.21 (0.66)
Self-regulated learning composite (1 to 7)	Posttest	4.52 (0.70)	4.29 (0.69)
Metacognitive self-regulation (1 to 7)	Baseline	4.11 (0.74)	4.10 (0.73)
Metacognitive self-regulation (1 to 7)	Posttest	4.47 (0.76)	4.22 (0.75)
Effort regulation (1 to 7)	Baseline	4.25 (0.71)	4.28 (0.70)
Effort regulation (1 to 7)	Posttest	4.55 (0.73)	4.41 (0.72)
Time and study environment (1 to 7)	Baseline	4.21 (0.72)	4.24 (0.71)
Time and study environment (1 to 7)	Posttest	4.54 (0.74)	4.33 (0.73)
Cognitive offloading tendency (1 to 7)	Baseline	3.62 (0.77)	3.64 (0.76)
Cognitive offloading tendency (1 to 7)	Posttest	4.43 (0.81)	3.82 (0.80)
Unit quiz (0 to 20)	Baseline	11.60 (3.00)	11.70 (2.90)
Unit quiz (0 to 20)	Posttest	14.70 (3.10)	14.20 (3.00)
Delayed retention quiz (0 to 20)	Follow-up	13.30 (3.20)	13.70 (3.10)

Table 3: Mixed effects model predicting posttest self-regulated learning composite.

Predictor	B	SE	95% CI	p
Intercept	0.02	0.06	[-0.10, 0.14]	.740
Baseline self-regulated learning	0.58	0.05	[0.48, 0.68]	< .001
Condition (GenAI)	0.24	0.09	[0.06, 0.42]	.010
School (private)	0.07	0.08	[-0.09, 0.23]	.390
Condition by school	0.10	0.10	[-0.10, 0.30]	.320
Stable home internet	0.12	0.06	[0.00, 0.24]	.048

Table 4: Mixed effects model predicting posttest cognitive offloading tendency.

Predictor	B	SE	95% CI	p
Intercept	-0.01	0.06	[-0.13, 0.11]	.880
Baseline offloading tendency	0.54	0.05	[0.44, 0.64]	< .001
Condition (GenAI)	0.46	0.08	[0.30, 0.62]	< .001
School (private)	0.05	0.08	[-0.11, 0.21]	.540
Condition by school	-0.06	0.10	[-0.26, 0.14]	.560
Stable home internet	0.09	0.06	[-0.03, 0.21]	.140

Table 5: Mixed effects model predicting delayed retention quiz score.

Predictor	B (points)	SE	95% CI	p
Intercept	0.42	0.58	[-0.73, 1.57]	.470
Baseline quiz score	0.61	0.06	[0.49, 0.73]	< .001
Posttest self-regulated learning	0.72	0.20	[0.33, 1.11]	< .001
Posttest offloading tendency	-0.58	0.19	[-0.95, -0.21]	.002
Condition (GenAI)	-0.32	0.28	[-0.87, 0.23]	.250
School (private)	0.44	0.30	[-0.15, 1.03]	.140

Table 4 reports the model predicting posttest cognitive offloading tendency. After adjustment, the generative AI condition was associated with higher offloading. The school and interaction terms were not statistically significant.

Table 5 reports the model predicting delayed retention. Baseline quiz score and posttest self-regulated learning were positively associated with delayed retention. Posttest cognitive offloading tendency was negatively associated with delayed retention after adjustment. The direct effect of condition on delayed retention was not statistically significant.

4. Discussion

The pattern observed in this study is consistent with the prevailing theoretical position that generative AI can strengthen learning when it is engineered into self-regulation cycles, yet it can also increase the propensity to offload cognitive work. The association between structured generative AI use and posttest self-regulated learning is interpretable within established self-regulation models that locate learning gains in frequent monitoring, strategic adjustment, and deliberate control of study behaviour [2-4]. When

learners are required to articulate goals, request explanations, verify outputs against instructional materials, and reflect on corrections, the tool may amplify metacognitive activity rather than displace it. This framing is compatible with current guidance that argues for instructional routines that render AI use auditable through process traces, verification, and revision [11,12].

The larger association between condition and cognitive offloading tendency is equally unsurprising given the nature of offloading. Offloading is a rational response to environments that reduce the cost of external support, and it is often adaptive when it redirects resources away from clerical demands [5]. However, offloading is also regulated by metamemory and effort costs, which means that tools that make externalisation effortless can shift self-regulation toward delegation rather than engagement [6,7]. Generative AI introduces a distinctive form of delegation because it can generate coherent reasoning paths and language that learners may adopt with minimal internal construction.

The negative association between offloading tendency and delayed retention aligns with a well-established learning science principle: durable learning depends on effortful retrieval and

elaboration, not merely on exposure to correct solutions [9,10]. If students routinely externalise the generative component of learning, they may reduce the very conditions that strengthen retention and transfer. The broader empirical literature is consistent with this interpretation. Evidence syntheses show positive mean effects of ChatGPT on learning outcomes alongside substantial heterogeneity, which is precisely what one would expect when tool use ranges from scaffolding to substitution [13]. Conceptual analyses likewise emphasise that AI can improve learning efficiency while simultaneously weakening engagement when the tool is used as a replacement for reasoning [14]. Recent experimental work has also raised concerns about cognitive under engagement during AI assisted writing, although the field will require further classroom studies with adolescents before strong causal claims are warranted [15].

The inclusion of both a public and a private school is salient for interpretation in the Philippine context because access conditions shape both uptake and function. Where feedback is scarce, structured generative AI may operate as a compensatory scaffold. Where resources are ample, it may drift toward convenience unless assessment practices reward explanation, verification, and revision. With only two schools, any school comparison must be interpreted contextually rather than as a population estimate, yet the design improves ecological validity relative to single site studies.

5. Conclusions

Generative AI is reshaping how students distribute cognitive effort across internal and external resources. When integrated through explicit planning, monitoring, verification, and reflection routines, generative AI use can align with stronger self-regulated learning. The same integration can increase cognitive offloading tendency, and higher offloading may coincide with weaker delayed retention when core generative work is delegated rather than practised. For senior high schools, the policy implication is not categorical prohibition, but careful instructional design that makes reasoning and verification compulsory and assessable so that AI functions as a scaffold rather than a substitute.

Acknowledgement

Conflict of Interest Statement

The authors declare no conflict of interest.

Author Contributions

Both authors have contributed equally. They have read and agreed to the published version of the manuscript.

References

[1] A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). Available online: https://files.eric.ed.gov/full text/ED338122.pdf (accessed on 02 December 2023).

[2] Zimmerman, B. J. Becoming a self regulated learner: An overview. *Theory Into Pract* **2002**, *41*, 64–70. https://doi.org/10.1207/S15430421TIP 4102_2

[3] Panadero, E. A review of self regulated learning: Six models and four directions for research. *Front Psychol* **2017**, *8*, 422. https://doi.org/ 10.3389/fpsyg.2017.00422

[4] Winne, P. H.; Hadwin, A. F. Studying as self regulated learning. In *Metacognition in Educational Theory and Practice*; Hacker, D. J., Dunlosky, J., Graesser, A. C., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1998; pp. 277–304.

[5] Risko, E. F.; Gilbert, S. J. Cognitive offloading. *Trends Cogn Sci* **2016**, *20*, 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[6] Hu, X.; Luo, L.; Fleming, S. M. A role for metamemory in cognitive offloading. *Cognition* **2019**, *193*, 104012. https://doi.org/10.1016/ j.cognition.2019.104012

[7] Gilbert, S. J.; Bird, A.; Carpenter, J. M.; Fleming, S. M.; Sachdeva, C.; Tsai, P. C. Optimal use of reminders: Metacognition, effort, and cognitive offloading. *J Exp Psychol Gen* **2020**, *149*, 501–517. https://doi.org/ 10.1037/xge0000652

[8] Sparrow, B.; Liu, J.; Wegner, D. M. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science* **2011**, *333*, 776–778. https://doi.org/10.1126/science.1207745

[9] Dunlosky, J.; Rawson, K. A.; Marsh, E. J.; Nathan, M. J.; Willingham, D. T. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychol Sci Public Interest* **2013**, *14*, 4–58. https://doi.org/10.1177/15291006 12453266

[10] Bjork, E. L.; Bjork, R. A. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*; Gernsbacher, M. A., Pew, R. W., Hough, L. M., Pomerantz, J. R., Eds.; Worth Publishers: New York, NY, USA, 2011; pp. 56–64.

[11] Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Geyer, C.; Göcks, J.; Hegelheimer, E.; Kasneci, G.; Khamis, M.; Krusche, S. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ* **2023**, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[12] Anders, A. D.; Speltz, E. D. Developing generative AI literacies through self regulated learning: A human centered approach. *Comput Educ Artif Intell* **2025**, *9*, 100482. https://doi.org/10.1016/j.caeai.2025.100482

[13] Wang, J.; Fan, W. The effect of ChatGPT on students' learning performance, learning perception, and higher order thinking: Insights from a meta analysis. *Humanit Soc Sci Commun* **2025**, *12*, 621. https://doi.org/10.1057/s41599-025-04787-y

[14] Jose, B.; Cherian, J.; Verghis, A. M.; Varghise, S. M.; Mumthas, S.; Joseph, S. The cognitive paradox of AI in education: Between enhancement and erosion. *Front Psychol* **2025**, *16*, 1550621. https://doi.org/10.3389/fps yg.2025.1550621

[15] Kosmyna, N.; Hauptmann, E.; Yuan, Y. T.; Situ, J.; Liao, X. H.; Beresnitzky, A. V.; Braunstein, I.; Maes, P. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv* **2025**. https://doi.org/10.48550/arXiv.2506.08872